Journal: Journal of the European Economic Association
Article doi: 10.1093/jeea/jvy036
Article title: Mitigating the Gender Gap in the Willingness to Compete: Evidence from a Randomized Field Experiment
First Author: Sule Alan
Corr. Author: Seda Ertac

INSTRUCTIONS

We encourage you to use Adobe's editing tools (please see the next page for instructions). If this is not possible, please list clearly in an e-mail. Please do not send corrections as track changed Word documents.

Changes should be corrections of typographical errors only. Changes that contradict journal style will not be made.

These proofs are for checking purposes only. They should not be considered as final publication format. The proof must not be used for any other purpose. In particular we request that you: do not post them on your personal/institutional web site, and do not print and distribute multiple copies. Neither excerpts nor all of the article should be included in other publications written or edited by yourself until the final version has been published and the full citation details are available. You will be sent these when the article is published.

1. **Licence to Publish:** Oxford Journals requires your agreement before publishing your article. If you haven't already completed this, please sign in with your My Account information and complete the online licence form. Details on how to do this can be found in the Welcome to Oxford Journals email.
2. **Permissions: Permission to reproduce any third party material in your paper should have been obtained prior to acceptance. If your paper contains figures or text that require permission to reproduce, please inform me immediately by email.**
3. **Author groups:** Please check that all names have been spelled correctly and appear in the correct order. Please also check that all initials are present. Please check that the author surnames (family name) have been correctly identified by a pink background. If this is incorrect, please identify the full surname of the relevant authors. Occasionally, the distinction between surnames and forenames can be ambiguous, and this is to ensure that the authors' full surnames and forenames are tagged correctly, for accurate indexing online.
4. **Figures:** If applicable, figures have been placed as close as possible to their first citation. Please check that they are complete and that the correct figure legend is present. Figures in the proof are low resolution versions that will be replaced with high resolution versions when the journal is printed.
5. **Missing elements:** Please check that the text is complete and that all figures, tables and their legends are included.
6. **Special characters and equations:** Please check that special characters, equations and units have been reproduced accurately.
7. **URLs:** Please check that all web addresses cited in the text, footnotes and reference list are up-to-date.
8. **Funding:** If applicable, any funding used while completing this work should be highlighted in the Acknowledgements section. Please ensure that you use the full official name of the funding body.
9. **Manuscript tagging:** The information in the table that appears before your manuscript contains the manuscript information that will be captured in the tagging online. Please check that this information is accurate as it cannot be changed after publication. If any of the information is incorrect, please mark the changes onto the table.

AUTHOR QUERIES - TO BE ANSWERED BY THE CORRESPONDING AUTHOR

The following queries have arisen during the typesetting of your manuscript. Please answer these queries by marking the required corrections at the appropriate point in the text.

Query No.	Nature of Query	Author's Response
Q1	Please note there is a charge of £350/€525/$600 per figure for colour reproduction of figures in print. Please confirm if you accept this charge. Alternatively, figures can be printed in black and white and appear in colour online only for no charge.	
Q2	Author: To check that we have your surnames correctly identified and tagged (e.g. for indexing), we have coloured pink the names that we have assumed are surnames. If any of these are wrong, please let us know so that we can amend the tagging.	
Q3	Author: Please provide JEL classification as per style.	
Q4	Author: Reference cited in Footnote 14 'Alan, Boneva, and Ertac (2016)' provide in Ref. list or delete otherwise.	
Q5	Author: Provide complete information of reference cited as 'Flory et al. (2014)' for Ref. List or delete otherwise.	
Q6	Author: Please provide title for Appendix section head.	
Q7	Author: Please update references Alan and Ertac (forthcoming), Gill et al. (forthcoming).	
Q8	Author: Please provide Institution name and Location in references Alan et al. (2017), Andersen et al. (2014), Dippel et al. (2017), Ertac and Szentes (2011), Kautz et al. (2014).	
Q9	Author: Please provide volume number in references Almlund et al. (2011), Croson and Gneezy (2009).	
Q10	Author: Please provide issue number in reference Blau and Kahn (2000).	

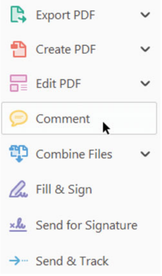
MAKING CORRECTIONS TO YOUR PROOF

These instructions show you how to mark changes or add notes to your proofs using Adobe Acrobat Professional versions 7 and onwards, or Adobe Reader DC. To check what version you are using go to **Help** then **About**. The latest version of Adobe Reader is available for free from get.adobe.com/reader.

DISPLAYING THE TOOLBARS

Adobe Reader DC

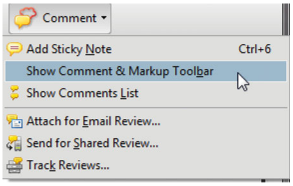
In Adobe Reader DC, the Comment toolbar can be found by clicking 'Comment' in the menu on the right-hand side of the page (shown below).



The toolbar shown below will then display along the top.

Acrobat Professional 7, 8, and 9

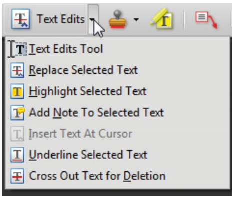
In Adobe Professional, the Comment toolbar can be found by clicking 'Comment(s)' in the top toolbar, and then clicking 'Show Comment & Markup Toolbar' (shown below).



The toolbar shown below will then be displayed along the top.

USING TEXT EDITS AND COMMENTS IN ACROBAT

This is the quickest, simplest and easiest method both to make corrections, and for your corrections to be transferred and checked.



1. Click **Text Edits**
2. Select the text to be annotated or place your cursor at the insertion point and start typing.
3. Click the **Text Edits** drop down arrow and select the required action.

You can also right click on selected text for a range of commenting options, or add sticky notes.

SAVING COMMENTS

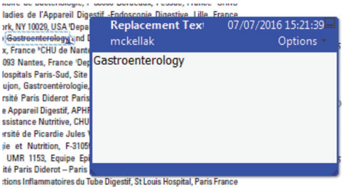
In order to save your comments and notes, you need to save the file (**File, Save**) when you close the document.

USING COMMENTING TOOLS IN ADOBE READER

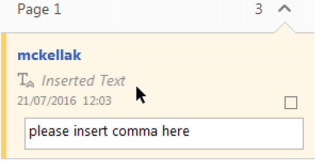
All commenting tools are displayed in the toolbar. You cannot use text edits, however you can still use highlighter, sticky notes, and a variety of insert/replace text options.

POP-UP NOTES

In both Reader and Acrobat, when you insert or edit text a pop-up box will appear. In **Acrobat** it looks like this:



In **Reader** it looks like this, and will appear in the right-hand pane:



DO NOT MAKE ANY EDITS DIRECTLY INTO THE TEXT, USE COMMENTING TOOLS ONLY.

Element	**Included in this paper**
Manuscript category	Article
Funder ID	10.13039/501100004412-Turkish Academy of Sciences, 10.13039/501100004412-Turkish Academy of Sciences
Orcid IDs	None
Contributors	Sule Alan, Seda Ertac
Supplementary data files cited	None
Related article(s)	
Corresponding author	Seda Ertac
Title	Mitigating the Gender Gap in the Willingness to Compete: Evidence from a Randomized Field Experiment
Keywords	None
Collection ID	

MITIGATING THE GENDER GAP IN THE WILLINGNESS TO COMPETE: EVIDENCE FROM A RANDOMIZED FIELD EXPERIMENT Q1

Sule Alan
University of Essex,
and Bilkent University

Seda Ertac
Koc University Q2

Abstract
We evaluate the impact on competitiveness of a randomized educational intervention that aims to foster grit, a skill that is highly predictive of achievement. The intervention is implemented in elementary schools, and we measure its impact using a dynamic competition task with interim performance feedback. We find that when children are exposed to a worldview that emphasizes the role of effort in achievement and encourages perseverance, the gender gap in the willingness to compete disappears. We show that the elimination of this gap implies significant efficiency gains. We also provide suggestive evidence on a plausible causal mechanism that runs through the positive impact of enhanced grit on girls' optimism about their future performance. (JEL:) Q3

1. Introduction

It is well-known that fewer women than men occupy top leadership positions in politics and the corporate world, and fewer women are represented in high-paying occupations that involve competitive paths; see Blau and Kahn (2000), Bertrand and Hallock (2001), Blau, Ferber, and Winkler (2002). Gender differences in attitudes toward competition have been put forward as an explanation for these findings; since if fewer women choose to compete, there will be less female winners in the competition for top positions, or in ambitious careers that usually involve competitive paths. In numerous lab and field settings women consistently exhibit a lower desire to compete and recent

The editor in charge of this paper was Imran Rasul.

Acknowledgments: We would like to thank the ING Bank of Turkey for providing funding. Ertac thanks the Turkish Academy of Sciences (TUBA-GEBIP program) and Alan thanks the ESRC Research Center on Micro-Social Change (MISOC) for financial support. We thank Thomas Crossley, Thomas Dohmen, Uri Gneezy, Jonathan Meer, and participants at the ASSA 2016 San Francisco meetings, MISOC Workshop, UCSD Rady School of Management strategy and economics seminar, and the ROA Conference in Maastricht University for comments and suggestions. We would also like to thank Elif Kubilay, Mert Gumren, Ipek Mumcu, Nergis Zaim, Banu Donmez, and Enes Duysak, as well as numerous other students who provided excellent research assistance. All errors are our own.

E-mail: salancrossley@gmail.com (Alan); sertac@ku.edu.tr (Ertac)

Journal of the European Economic Association 2018 00(0):1–39 DOI: 10.1093/jeea/jvy036

studies document the strong link between this attitude and actual choices/outcomes in education and the labor market; see Flory, Leibbrandt, and List (2015), Reuben, Sapienza, and Zingales (2015), and Buser, Niederle, and Oosterbeek (2014).[1] From an economic standpoint, an important concern here is efficiency: if males and females are equally able in tackling a task and if females shy away from competition involving such a task, then winners of tournaments will on average be less able than if males and females had similar entry rates.

A number of studies use laboratory experiments to explore the effectiveness of policies that aim to mitigate gender differences in competitiveness. For example, one strand of the literature considers affirmative action and preferential treatment through changes in tournament rules favoring women; see Balafoutas and Sutter (2012), Niederle, Segal, and Vesterlund (2013), and Sutter et al. (2016). These studies find that women enter tournaments more frequently with such policies, without sacrificing efficiency. Booth and Nolen (2012) show that single-sex schooling might eliminate gender differences in competitiveness, whereas Petrie and Segal (2014) find that the gender difference disappears if tournament prizes are high enough.[2] Implicit in these studies is that competitiveness may not be a fixed trait and is likely to respond to external and environmental factors. Consistently with this, Almas et al. (2016) highlight the role of family background and socioeconomic status in determining competitive behavior. Gneezy et al. (2009) and Andersen et al. (2013) show that social structure may influence competitiveness, with gender gaps being nonexistent in matrilineal societies with more gender equality.

In this paper, we approach competitiveness from a different angle. Underlying this approach is the conjecture that competitiveness is linked to a set of noncognitive skills related to drive and motivation in performance contexts. In particular, we conjecture that competitive behavior, especially in a dynamic setting, is related to grit, a noncognitive skill that involves challenge-seeking and passion for long-term goals, and is closely linked to perseverance and tenacity. What distinguishes a gritty individual is her beliefs regarding the role of effort in the performance process and her response to performance feedback. A gritty individual tends to set ambitious performance goals with the belief that these goals are attainable through persistent effort, and tends to persevere after initial failure; see Duckworth et al. (2007) and Duckworth and Quinn (2009). Based on item-set inventories developed by Angela Duckworth and colleagues, grit has been shown to predict achievement in various educational and occupational settings; see Duckworth et al. (2011), Eskreis-Winkler et al. (2014), and Maddi et al. (2012).

1. Bertrand (2011) and Croson and Gneezy (2009) provide in-depth reviews of gender differences in economic preferences and psychological attitudes. For the gender gap in the response to competition and competitiveness, see Gneezy et al. (2003), Niederle and Vesterlund (2007, 2011), and the references therein. There is also evidence showing that the gender-competition gap may emerge quite early in the life cycle and persist into adulthood (Sutter and Glatzle-Rutzler 2015), as part of a growing literature studying competitiveness in children and adolescents in different cultures; see Andersen et al. (2013), Cárdenas et al. (2012), Dreber, Von Essen, and Ranehill (2014), and Khachatryan et al. (2015).

2. Lee et al. (2014), on the other hand, find that single-sex schooling has no effect on the gender gap.

The objective of the current paper is to explore whether competitiveness and the gender gaps therein are responsive to an intervention that fosters grit. To this end, we evaluate a unique, large-scale educational program targeted at elementary school children in Turkey. Implemented by children's own (trained) teachers within the classroom environment, the program involves covering a carefully designed curriculum that highlights the role of effort in achieving goals, encourages challenge seeking, and promotes a constructive response to performance feedback. We provide a brief review of the content of the curriculum in Section 3.

At the design stage of the field experiment, we hypothesized that in addition to its effects on achievement outcomes, the intervention may have an impact on attitudes toward competition and the gender gaps therein. The rationale behind this hypothesis is that someone who internalizes the ideas advocated in the intervention and realizes that ambitious goal-setting and perseverance have higher returns than previously thought may have a changed attitude toward competition. Speaking to the well-established literature on the gender gap in competitiveness, the current paper constitutes a formal test of this hypothesis. Although the program does not have a gender focus at all, the ideas promoted by the program may be particularly effective for girls, who tend to be more pessimistic about their performance, and therefore tend to make sub-optimal choices by shying away from tournaments (Niederle and Vesterlund 2007). Given this, we are interested in understanding whether the educational intervention has heterogeneous effects with respect to gender in a way that can mitigate the gender-competition gap, and improves girls' payoffs from an efficiency standpoint.

The intervention is designed as a randomized-controlled trial and implemented twice in this manner using two independent samples of schools. Each study follows the same randomization procedure and uses the same education material, providing us with two large independent samples to estimate the treatment effect on the willingness to compete. Our outcome measures come from an incentivized mathematical real-effort task, whereby children choose to compete, receive performance feedback, then make a choice again. This dynamic nature of our experimental task is intended to capture a realistic performance setting where rather than making one-time, permanent choices, individuals observe how they fared, interpret the feedback they receive, and revise their decisions.

We first document that in our control sample, there is a statistically significant 7.8 percentage point gender gap in the willingness to compete in the first stage of the competition task. Despite the fact that there is no gender difference in actual performance and therefore in the probability of receiving negative performance feedback, this gap is estimated to be 9.6 percentage points in the second stage competition, after feedback. The size of the gap appears to be independent of the type of feedback received in the first stage.

We then show that the gender gap that exists before and persists after feedback is closed by the treatment. We estimate a statistically significant treatment effect on the willingness to compete for both boys and girls in the first stage, with the effect on boys being somewhat weaker. The effect is large and significant for girls, closing the first stage gender gap. Moreover, we document a statistically significant treatment effect on competitiveness after feedback only for girls, entirely eliminating the second stage gender gap. The effect on girls' competitiveness is of considerable size: the propensity to compete in the second stage is about 15 percentage points higher for girls in treated schools compared to girls in untreated schools. We also show that the higher willingness to compete induced by the treatment brings significantly higher rewards to girls in treated schools relative to the girls in the control group. This improved efficiency stems from the fact that the treatment lowers the propensity of girls to shy away from competition. Our rich data allow us to contemplate a potential causal mechanism that points to a significant increase in optimism about future performance due to enhanced grit, to explain these results.[3]

The paper advances a large literature that documents a pervasive gender gap in competitiveness by showing, for the first time, that competitiveness is a malleable trait and the inefficient gender-competition gap can be eliminated early in childhood. Exposing children to a positive worldview that highlights the role of effort in achieving challenging goals and encourages constructive interpretation of performance feedback can go a long way toward achieving this end. The unique educational program we evaluate in the paper is an example of an intervention that can successfully impart this worldview and can be cost-effectively implemented in the classroom environment. Finally, highlighting the importance of studying gender gaps in competitiveness in a dynamic context, our post-feedback results complement the recent literature on performance feedback and its effects on choices and future performance.[4]

The rest of the paper is structured as follows: Section 2 provides background information on the program we evaluate and presents the evaluation design, Section 3 describes the content of the intervention, Section 4 describes our experimental outcome measures, Section 5 presents the results and a mediation analysis, and Section 6 concludes.

2. Program Background and Evaluation Design

Our study sample is drawn from state-run elementary schools in relatively deprived areas of Istanbul, representing Turkey's lower socioeconomic segment. The Turkish Ministry of Education encourages schools and teachers to participate in socially useful extra-curricular programs offered by the private sector, NGOs, the government and international organizations. All elementary school teachers are given a maximum of 5 hours per week to be involved in these programs. Their participation is voluntary and if they choose not to participate in any program, there is no restriction on the way in which these hours are used. The program we evaluate in this paper is implemented as an

3. 2012 PISA results reveal that countries with large gender gaps in perseverance tend to have larger gender gaps in mathematics performance (OECD 2013).

4. See Azmat and Iriberri (2010), Barankay (2011), Eriksson et al. (2009), Ertac and Szentes (2011), Gill and Prowse (2014), Gill et al. (forthcoming), and Wozniak, Harbaugh, and Mayr (2014).

extra-curricular project, under the oversight of the Education Directorate of Istanbul. The main objective of the program is to improve key noncognitive skills in elementary school children in the classroom environment by training their teachers.[5]

The program was designed in Spring 2013 as a randomized-controlled trial and implemented twice in this manner, using two independent samples of elementary schools across Istanbul. Our first sample originally contains 37 schools. We randomly allocated 15 schools to initial treatment (IT), 10 schools to control-then-treatment (CT) and the remaining 12 schools to pure control (PC). For this sample the intervention had two main treatment arms, each of which had a specific behavioral target. The first arm aimed to improve the ability to make decisions in a forward-looking manner and encourage patience. The evaluation of this arm shows that the intervention leads to more patient intertemporal choices and improves behavioral conduct in school; see Alan and Ertac (forthcoming).[6] The second arm, which was implemented in the same 15 IT schools after the implementation of the first arm, aimed to foster grit. The effect of this arm was first evaluated within a novel real effort task that aims to measure several aspects of grit, and with respect to objective test scores. Results show that children who receive the educational intervention on grit/mindset can actually achieve success and higher payoffs in the real-effort task, when given the time to work/practice for it. We also find that the intervention improves math test scores; see Alan, Boneva and Ertac (2017).

After we collected our grit measure and test scores, a random subset of these schools (6 IT, 3 CT 4 PC, total of 13 schools) were exposed to another treatment (a gender-role model exposure intervention) before we collected our competition task data in May 2014, therefore we remove those schools from our current analysis, which leaves us with 23 schools (approximately 1300 students).[7] With this design, we can evaluate the independent impact of the patience treatment on our outcome measures, and indeed show that the patience treatment by itself does not have any impact on competitiveness (see Table B.1 in the Online Appendix). Still, we would not be able to isolate or rule out potential complementarities across grit and patience training using this study sample. Such an issue does not arise in the second study.

In the second study, we randomly assign only the grit training arm of the first study across a new set of schools in Istanbul. The intervention follows the same procedures, with the same curricular materials and the same teacher training approach. This sample consists of 16 schools (8 treatment, 8 control) and has a total of about 1,300 students.[8]

5. A growing body of research shows that certain "noncognitive skills" are strongly associated with achievement in various economic and social domains. A strand of this literature focuses exclusively on children, and shows that childhood period is crucial for development of these skills; see Almlund et al. (2011), Borghans et al. (2008), Heckman et al. (2006, 2010, 2014), and Kautz et al. (2014).

6. We lost 1 CT school, which had only a single classroom, after the first follow up due to an involuntary teacher rotation.

7. Including these schools in our control group does not materially alter our main findings; results are available upon request.

8. In the second study, more teachers per school stated their willingness to participate in the program relative to the first study, giving us a smaller number of schools with approximately the same sample size as the first sample.

In both studies, the randomization was performed in the following way. First, the Istanbul Directorate of Education sent the official documentation of the program to all elementary schools in chosen districts of Istanbul.[9] The teachers in these schools were then contacted in random sequence and invited to participate in the program. Teachers were informed that upon participation they would be assigned to different training phases within the coming two academic years. All teachers who agreed to participate were promised to eventually participate in training seminars and receive all training materials, but they were not told when within the next two academic years they would receive the treatment, until the random assignment was completed. The promise of the training offer was made to the teacher and not to current students, that is, although children in control groups never received the training as they moved on to middle school after 4th grade, their teachers did, albeit at a later time. This was done in order to allow for long-term follow-up.

Once a teacher stated a willingness to participate, we randomly assigned their school into the treatment or the control group. Note that in both studies the unit of randomization was the school, not the classroom, in order to prevent potential spillovers across classrooms. In the first study, a given school where there was a teacher willing to participate had a 40% ex ante chance of being assigned to the initial treatment group (patience + grit), a 30% chance to be in the second treatment group (patience only) and 30% chance to be assigned to the control group. In the second study, a given school where there was a teacher willing to participate had a 50% chance of being in the treatment group. For each study, we stopped recruiting teachers when we hit the logistical constraint of being able to physically visit the classrooms.[10] In the first study, baseline data were collected in Spring 2013, the first intervention (patience) was implemented on 15 IT schools in Spring 2013, and the intervention on grit was implemented in Fall 2013 on the same 15 schools. The follow-up data on competition task were collected in May 2014. In the second sample, baseline data were collected in Spring 2015, the intervention (grit only) was implemented in Fall 2015, and the follow-up data were collected in January 2016. Full details of the evaluation design, including the timeline of the intervention and measurement phases for each study sample, are given in Table A.[11]

The sample generated with this design contains schools in which at least one teacher stated their willingness to participate in the program.[12] Therefore, the estimated impact

9. The program was titled "financial literacy, savings and economic decisions" and no further information on the particulars of the program were disclosed to the teachers prior to the teacher training seminars.

10. In order to ensure data quality, the authors coordinated the field logistics, trained a select group of students and experienced interviewers to assist with data collection, and physically visited all classrooms to implement the tasks and collect data. All measurements were conducted with the approval of the local IRB and the permission of the Ministry.

11. We do have long-term follow up results on test scores as well; see Alan, Boneva, and Ertac (2017).

12. Participating in the program is not a matter of choice for children as this is a program that is approved by the Ministry of Education and the school principal. Once a teacher participates, all the students in his/her class participate. Similarly, all students in the classes of teachers who volunteered but were assigned to the control group participate in our measurements, and we have their data.

of the program is the average treatment effect on the treated and in principle, is not readily generalizable to the population. However, in the first study, approximately 60% of the contacted teachers accepted our offer and the most common reason for nonparticipation was being “busy with other projects, although happy to participate in this program at a later date” (about 20%). The rest of the nonparticipation was due to “impending transfer to a school in another city, with a willingness to participate if the program is implemented there” (about 5%), and “not being in a position to participate due to private circumstances” (about 10%). In study sample 2, acceptance of the training offer reached 80%. Given these numbers, the external validity of our results is likely quite strong and our estimated average treatment effects on the treated are very close to population treatment effects. We should also note that it is quite common for elementary school teachers to participate in officially approved extracurricular projects offered by reputable organizations. Therefore, children in the control group also participate in extracurricular projects often covering topics such as health, the environment, consumer responsibility, and citizenship.

In terms of implementation of the program, although we do not track teachers' activities, we did an anonymous ex post survey with the teachers in sample 1 on how intensely they implemented the materials. Of the 28 teachers who answered this survey, 25% implemented the training very intensely, 68% implemented it moderately intensely whereas 7% (only 2 teachers) report that they did not have much chance to implement the materials.

After establishing that the two samples represent the same population using the baseline characteristics, we pool them to increase the number of clusters (number of schools) and define a treated school as one where teachers received the grit treatment, either as “grit only” (as in study 2) or as “grit plus patience” (as in study 1). With this, we have a total of 17 schools in the treatment group and 22 schools in control, giving us about 2600 students in total. We also provide the main results separately for each sample and a related discussion in the results section.

3. The Educational Intervention on Grit

The educational initiative required the production of a rich set of educational materials, which involved a broad interdisciplinary endeavor. Although the target concepts of the materials were determined by the authors, specific contents (e.g., scripts) were shaped with input from an interdisciplinary team of education psychologists, a voluntary group of elementary school teachers, children's story writers, and media animation artists, according to the age and cognitive capacity of the students.

The program involves providing animated videos, mini case studies, and classroom activities that highlight (i) the plasticity of the human brain against the notion of innately fixed ability, (ii) the role of effort in enhancing skills and achieving goals, (iii) the importance of a constructive interpretation of setbacks, failures and success, and (iv) the importance of goal setting. The aim of the training program is to expose students to an optimistic world view in which any one of them can set goals in an area of their interest and can work toward these goals by exerting effort. The materials highlight the idea that in order to achieve these goals, it is imperative to avoid interpreting immediate failures as a lack of innate ability or intelligence. This world view encompasses any productive area of interest, whether it be music, art, science or sports.[13] Visual materials and stories are supplemented by classroom activities created and supervised by teachers, based on general suggestions and guidelines put forward in teacher training. It should be noted that there is no mention of competitiveness or emphasis on doing better than others in the materials, and success in the examples given to children means reaching individualistic goals, such as mastering a task that one finds challenging.

Weekly topics, main materials and supplementary activities are very clearly defined and specific guidelines on how to structure each lesson are prepared for the teachers (as part of a teacher kit). However, the program is not merely a set of materials to be covered in a specified period of time, like a regular curriculum item. Instead, it aims to change students' beliefs about the role of effort in performance processes and the potential return to perseverance partly by changing the mindset of the teachers and the nature of the classroom environment. To this end, in addition to covering the curricular items as suggested, teachers were strongly encouraged to adopt a teaching philosophy that emphasizes the role of effort in everyday classroom tasks, for example, while giving performance feedback and interpreting test results.[14]

It is also important to note that the intervention is not directed at girls at all. Rather, it aims to implement, in the minds of both girls and boys, the idea that effort has a paramount role in success and that there is a return to perseverance after failure, if one works hard. The intervention also has no direct material on self-confidence or over/under-confidence, and does not aim to make children more confident, in the sense of believing that they are better at a task than others. Still, the kind of positive worldview on the productivity of effort that the program advocates is expected to encourage a gritty attitude toward challenging tasks despite initial failure, and optimism about future performance. If one believes that effort will pay off, they are going to be more Q4
willing to take on challenges and ambitious tasks, and will be less likely to give up once they fail, as they are more likely to believe that they can improve upon their previous performance. Based on this conjecture, we designed our two-stage competition task, which we describe in what follows.

13. More information about the curriculum, with all covered topics and classroom activities can be found in the Online Appendix.

14. Alan, Boneva, and Ertac (2016) show that this intervention is highly effective in inducing ambitious goal-setting, perseverance and individual skill accumulation. Students in treated schools also obtain significantly higher scores in standardized math and verbal tests, without any heterogeneity in any of the effects with respect to gender.

4. The Outcome Measure: A Two-Stage Competition Task

Our outcome measure is designed to estimate the effect of the intervention on initial competitiveness and competitiveness in response to performance feedback. The task we use is an addition task, which involves adding two 2-digit numbers and one single-digit number. The task consists of three performance periods and children are given 2.5 min in each period. One of these periods is selected randomly at the end, and rewards are determined based on the performance and decisions in the selected period. For all three periods, each student is matched with another student from a different school (whom we will call "opponent" hereafter), who had done the same tasks before and whose performance was recorded.[15]

Extending the widely used competition task in the literature (Niederle and Vesterlund 2007), where individuals choose which incentive scheme to work under, our design adds one more choice stage to elicit competitive behavior after feedback.[16] In the first period (which we call period 0 here for ease of reference), students perform the addition task under a piece-rate incentive scheme, whereby they receive 1 token for every addition they are able to do correctly. Each token corresponds to a gift from a basket that includes items of value to children, such as attractive toys and stationary. In period 1 (the first competition choice stage), students have a choice between piece-rate and competing with their matched opponent. If they choose the piece-rate, they are rewarded with one token per correct answer. If they choose the tournament, they are rewarded with 3 tokens per correct answer, but only if their performance exceeds that of the opponent. In the case of having a lower performance, a student that chooses to compete receives zero tokens, whereas in the case of a tie, she gets 1 token per correct answer. We inform the students that choosing competition means that their performances will be compared with the pre-recorded performance of a child in another school, who already did the task. We make sure to emphasize that the children will not meet their opponent and the competition is not real-time, and we just use the opponent's performance in this task as a benchmark if the child chooses to compete. We are not explicit about the incentive scheme faced by the opponent when he/she performed; in particular, we say nothing that suggests the opponent made a choice (this was also never asked by the students). The fact that children in both treated and control schools have the same information about the opponent reassures us that the treatment effects we estimate do not stem from this aspect of the design.[17]

15. We implemented the same tasks in the same order in three classrooms in three pilot schools, and recorded repeat performances with a piece-rate. Each student in our sample is randomly matched with one student from this "benchmark sample".

16. Designs with two competition choice stages have also been utilized in Andersen et al. (2014) and Berlin and Dargnies (2016).

17. This is also corroborated in an analysis of children's expectations about the opponent's performance—children in treated and untreated schools are no different in terms of how well they expect the opponent to do, in either the first or the second stage (p-value for stage 1 is 0.99 and for stage 2 is 0.91).

After children choose the incentive scheme, we elicit beliefs/expectations regarding their performance in the next stage. Specifically, children are asked to state their beliefs about (1) the number of correct answers they will have, (2) the number of correct answers of their opponent. These guesses are incentivized in the following way: Children are told that three people in their class will be randomly chosen after the experiment ends, and these three will get an extra small gift for each correct guess. Children are also reminded that there is always an incentive to do as many additions as they can in the actual task and it does not make sense to stop just to be consistent with their guess. Using these guesses, we capture optimism about future relative performance, which directly relates to the educational material in the program. After performing the task in period 1, children receive feedback about (1) the number of additions they were able to do correctly, (2) whether their performance was better than, worse than, or equal to that of the opponent. In period 2 (the second competition choice stage), children make a choice again between piece-rate and competition against the same opponent, to be implemented if that period is chosen for payment. After making this choice, they state beliefs again about their own and opponent's performance in the upcoming stage, in the same way as in the first choice stage. This allows us to understand how children form beliefs about future performance in response to performance feedback. If the child chooses to compete in any period, her results are compared with the performance of the opponent in the corresponding period, and rewards are determined based on this comparison.

Children learn the gender of their (randomly assigned) opponent before making their initial competition choice. That is, they know whether they are matched with a boy or a girl. Before this, there is also an incentivized question that elicits gender stereotypes: children are told about one student (with a girl's name) and another (with a boy's name), randomly selected from a different school/class, and are asked to guess which one has done better in this same task. If they guess who actually did better correctly, they get a small extra gift.

A summary of the timeline from the perspective of the participant is given in what follows.

- Period 0: Do the task under piece-rate
- Period 1 (competition choice, stage 1):
  - Guess which child (the one with a girl's name or the one with a boy's name) would do better in this task
  - Learn about opponent's gender
  - Choose between piece-rate and competition
  - Make guesses about own and opponent's performance
  - Do the stage 1 task under chosen incentive scheme
  - Receive feedback: Learn own performance and whether it was better, equal or worse than the opponent's
- Period 2 (competition choice, stage 2):

– Choose between piece-rate and competition
– Make guesses about own and opponent's performance
– Do the stage 2 task under chosen incentive scheme

To implement the previous design, children were given workbooks, and were instructed to turn over pages only at specified times. Each child's workbook had a code that matched it to a pair of actual period 1 and period 2 performances that came from a student in one of the pilot classes. The workbooks were randomly distributed, rendering a random match in terms of opponent gender and opponent performance. After the task was completed in period 1 (the first choice stage), feedback was given to children by experimenters by writing the child's own actual performance and circling the outcome of the performance comparison (the child having a worse, better or equal performance compared to the matched child) on the sheet. Instructions are provided in the Appendix.[18]

It is worthwhile to re-emphasize here that the intervention itself did not involve any material related to competition, competitiveness or doing better than others, and had an individualistic focus on skill development and performance improvement.[19] In this sense, for the children, the competition task was not something that is directly and easily relatable to the education they received, and since we conducted a set of unrelated experiments/surveys, it was unlikely for them to associate us experimenters with the educational program either. We also took additional measures to prevent potential demand effects—teachers were not in the classroom during the experiments, they did not have any information about our measurements, and we made sure to set strong incentives by using gift items that are very attractive for children of this age and socioeconomic group.

5. Results

We have data from approximately 2,600 students in 39 schools. About 51% of our sample is male, and the average age is about 9.5 years, as is typically the case for 4th graders in Turkey. We collected baseline information on students using official records

18. One concern here might be about the potential effects on subsequent performance of asking whether a boy/girl is better or providing information on opponent gender. However, since this information/questions are provided to both treatment and control groups in the same way (opponent gender is written in a box in the choice sheet), any effects they may have are expected to be balanced. Consistently with this, treatment effects display no heterogeneity across opponent gender (related results are available upon request).

19. Because choosing to compete is a higher-reward performance goal in the specific context of our task than choosing the individualistic incentive scheme, we can expect gritty individuals to be more likely to choose to compete in our task, before and after feedback. We should note, however, that this does not mean grit is less relevant under individual performance incentives. In noncompetitive contexts as well, gritty individuals can set ambitious performance goals and work toward them perseverantly (e.g., choose a challenging performance goal versus a less challenging goal, and sticking with a challenging goal after initial failure), as reflected in the intervention material, which involves individualistic rather than competitive contexts.

TABLE 1. Randomization balance.

	Full sample		Girls		Boys	
	Control	Difference	Control	Difference	Control	Difference
Panel 1: Baseline variables						
Age	9.52	0.02	9.48	0.04	9.54	0.01
		(0.03)		(0.04)		(0.02)
Raven score (normalized)	0.05	−0.10	0.08	−0.07	0.02	−0.14
		(0.13)		(0.13)		(0.14)
Math grades (normalized)	0.08	−0.17	0.14	−0.12	0.02	−0.21
		(0.15)		(0.13)		(0.17)
Verbal grades (normalized)	0.13	−0.27	0.29	−0.23	−0.02	−0.30
		(0.19)		(0.18)		(0.22)
Risk tolerance	2.29	0.11	2.17	0.18	2.40	0.05
		(0.174)		(0.17)		(0.14)
Male teacher	0.26	−0.09	0.25	−0.08	0.26	−0.09
		(0.09)		(0.08)		(0.09)
Wealth (teacher-reported 1–5)	2.72	0.07	2.70	0.13	2.74	0.01
		(0.16)		(0.19)		(0.14)
Grit score (normalized)	0.03	−0.07	0.23	−0.10	−0.15	−0.05
		(0.11)		(0.13)		(0.11)
Panel 2: Within-task variables						
Correct answers in piece-rate stage	5.78	−0.29	5.90	−0.13	5.66	−0.45
		(0.39)		(0.40)		(0.42)
Negative feedback in stage 1	0.42	0.04	0.41	0.02	0.44	0.06
		(0.05)		(0.05)		(0.06)
Male opponent	0.50	−0.01	0.50	−0.02	0.50	−0.00
		(0.02)		(0.03)		(0.03)

Notes: Each row reports coefficients from a regression of the variable shown in the first column on the treatment dummy. The first column in each group reports the mean of the control, the second one reports the difference between the treatment and control. Panel 1 presents the balance for demographic variables and baseline attitudes either reported by the child or the teacher. The variable "Grit score" is an extracted factor from questionnaire items in the pretreatment student survey. The Raven score is measured using a progressive Raven's matrices test (Raven, Raven, and Court 2004). Risk tolerance is elicited using the incentivized Gneezy and Potters (1997) task. Students' family wealth is reported by the teacher (1–5 scale). Math and verbal grades, obtained prior to the implementation of the program, are official grades given by teachers. Panel 2 presents the balance for (i) the performance in the first period, (ii) the proportion of negative feedback in stage 1, and (iii) the proportion of male opponents. Standard errors, obtained via clustering at the school level, are reported in parentheses.

(for grades), surveys, cognition tests, and experimental elicitation tasks. Although some survey information come directly from children, some are obtained via teacher assessment surveys. These variables allow us to test the balance of the treatment groups in our data and provide us with useful covariates that are predictive of our outcome measures.

5.1. Internal Validity

We first check whether our pooled data are balanced across treatment status with respect to a number of student characteristics, collected at the baseline stage. Table 1 shows results from ordinary least squares regressions of baseline variables on the treatment dummy for the full sample, as well as for boys and girls separately. Although the columns labeled as "control" give the mean of the control for the respective variable and the respective group, the columns labeled as "difference" show the difference between control and treatment. Panel 1 presents the balance results for demographic variables and baseline behaviors and attitudes, and panel 2 presents the balance results pertaining to three variables that come from our competition task.

The variables in panel 1 are constructed as follows: Age is reported by the student, whereas student's family wealth is reported by the teacher using a 5-point scale. Cognitive ability is measured using Raven's progressive matrices test (Raven, Raven, and Court 2004) and risk tolerance is measured using a task based on Gneezy and Potters (1997).[20] Math and verbal (Turkish) grades are official (pretreatment) end-of-semester grades. The grit score is a standardized factor extracted using eight item-set questions adapted from the "short grit scale" in Duckworth and Quinn (2009). The questions mainly aim to elicit self-reports on gritty/nongritty behaviors on the part of the children (such as their attitude toward difficult tasks, how hard they work, how discouraged they get after a failure; see the Appendix for the full set of questions). We collect this measure prior to the implementation of the program as well as in the follow-up.

As can be seen from the table, differences are not statistically different from zero across treatment and control for any of the variables for any subgroup. This assures us that the pooled data are balanced across treatment status conditional on gender as well as for the whole sample, and our results are internally valid. This balance result is not surprising, as our school sample is quite homogeneous, all of them being state schools in relatively underprivileged areas of Istanbul (the target areas when designing the intervention were these relatively lower socioeconomic-status areas, which was also desired by the Directorate of Education). Reflecting this homogeneity, the average SES (family wealth as reported by the teacher) per school is balanced across treatment groups ($p = 0.98$), and there are no significant differences in class size ($p = 0.64$) or the age/gender of the teacher ($p = 0.59/0.88$).

Panel 2 of Table 1 presents the balance of in-task variables. An important finding to note here is that there is no difference in piece-rate performance across our treatment groups. Nor is there any difference in the probability of receiving negative feedback after the first stage. We will refer to these findings when we discuss the gender gaps in competitiveness we observe in the baseline in the next section.

20. In this task, children have an endowment of 5 tokens, which they can allocate between a riskless and a risky option. Tokens allocated into the risky option (conveyed to children as putting tokens in a bowl), are tripled with 50% chance and lost with 50% chance, based on the color of a ball drawn from an opaque bag. Tokens that are not allocated into the risky option are safe. The number of tokens placed into the risky option is a measure of risk tolerance.

TABLE 2. Gender gap in competition choice (control sample).

	Stage 1		Stage 2	
	(1) No covariates	(2) With covariates	(3) No covariates	(4) With covariates
Female	−0.078***	−0.075***	−0.096***	−0.089**
	(0.02)	(0.02)	(0.04)	(0.04)
Male opponent		0.027		0.008
		(0.02)		(0.03)
Risk tolerance		0.042***		0.025***
		(0.01)		(0.01)
Raven score		0.052***		0.039***
		(0.01)		(0.01)
Male teacher		−0.055		−0.042
		(0.04)		(0.03)
Baseline grit score		0.024*		0.047***
		(0.01)		(0.02)
Optimism (Relative P.) S1		0.070***		
		(0.01)		
Optimism (Relative P.) S2				0.130***
				(0.02)
Observations	1354	1154	1351	1154

Notes: This analysis is based on the control sample only. Coefficient estimates are average marginal effects from a logit regression where the dependent variable is the binary outcome of competition choice. The first row gives the estimated percentage point gender gap in the willingness to compete. The variables Optimism (Relative P.) S1 and S2 are constructed by taking (and normalizing) the difference between own expected performance minus the expectation about the opponent's performance for stage 1 and stage 2, respectively. Clustered standard errors (at the school level) are in parentheses. $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

5.2. Willingness to Compete and Gender Gaps in the Control Group

We now turn our attention to the gender gap in the baseline. Focusing on elementary school children and using a dynamic version of a well-known experimental task, this initial analysis provides new evidence on the prevalence of the gender gap in the willingness to compete and gives us the baseline gender gap figures upon which our treatment operates. Table 2 presents average marginal effects from logit regressions where the dependent variables are the binary competition choices in stage 1 and stage 2. The first point to note in this table is that there is a statistically significant (unconditional) gender gap between girls and boys in the first stage, with girls 7.8 percentage points less likely to choose to compete (see column (1)). The gender gap remains, and in fact slightly widens in the second stage, after feedback: We estimate that girls' willingness to compete is now 9.6 percentage-points lower than boys, although the difference in the gap between the two stages is not statistically significant (p-value = 0.68).

A major question that the gender-competition literature has focused on is the determinants of competitiveness, which also provides insights about the nature of the gender gap. Columns (2) and (4) of Table 2 include the main potential determinants of competition choice as explanatory variables: beliefs about future performance and risk tolerance, in addition to cognitive ability (Raven score), opponent gender, and teacher gender. We add to these the normalized baseline grit score to show the relationship between competitiveness and grit at the baseline.[21] A major factor that influences competition choices is beliefs on how likely one is to do well in this task. Our dynamic competition task allows us to construct a variable that captures such beliefs before and after feedback. This measure, which is available for each choice stage, is constructed as the difference between the child's (incentivized) guess about her own performance and her guess about her opponent's performance in the upcoming stage, which we refer to as "optimism" about relative performance from here on. Although this measure is undoubtedly related to self-confidence about relative ability, it also involves expectations of future effort, which are particularly important in our context because the intervention promotes an optimistic view about future success provided that one works hard. In that sense, this measure is different than the belief (self-confidence) measure commonly used in the extant competition literature that involves guessing one's rank relative to others based on a past, realized performance, which is why we prefer to call it optimism. For example, a child who is not particularly self-confident about a realized past performance relative to others may still be optimistic about her future performance, if she thinks that she will need to put more effort next time and believes that such effort will bring success, as the intervention material emphasizes.[22]

As expected, higher optimism about performance relative to the opponent, risk tolerance and cognitive ability are associated with a higher propensity to compete in both stages, with optimism having the highest predictive power: a one standard deviation increase in expectations toward more optimism increases the probability of choosing to compete by 7 percentage points in the first stage and about 13 percentage points in the second stage. Note that opponent gender seems to have no effect on the willingness to compete in either stage. We find that the grit score is also predictive of competition choice in both stages. The predictive power of the grit score in the second stage is of considerable size, rivaling that of risk tolerance and cognitive score: a one standard deviation increase in the grit score is associated with a 5 percentage-point increase in the willingness to compete after feedback, when optimism about future performance is controlled for. This relationship between competitive choices and grit provides the basis for investigating the effect of grit training on competitiveness, the main motivation of the paper.

Our data also provides evidence that competitiveness in a dynamic context (persistence in a competitive path in particular) is related to actual educational outcomes as measured by mathematics grades, even quite early on. Table 3 presents the predictive

21. What the experimental task elicits using incentive scheme choice is a composite measure that potentially involves beliefs, risk tolerance, and the pure taste/distaste for competition.

22. All continuous variables, with the exception of risk tolerance that contains the number of tokens allocated into the risky option in the Gneezy and Potters investment task, are normalized to facilitate the interpretation of the coefficient estimates.

TABLE 3. Competition choice and mathematics grades.

	Stage 1				Stage 2			
	Girls	Girls	Boys	Boys	Girls	Girls	Boys	Boys
Competition choice in S1	−0.027	−0.092	0.190**	0.048				
	(0.08)	(0.06)	(0.07)	(0.06)				
Grit score		0.346***		0.331***		0.320***		0.301***
		(0.04)		(0.05)		(0.04)		(0.04)
Optimism (Relative P.) S1		0.075*		0.057**				
		(0.04)		(0.02)				
Raven score		0.256***		0.281***		0.249***		0.275***
		(0.04)		(0.05)		(0.04)		(0.05)
Risk tolerance		−0.078*		0.005		−0.088**		−0.004
		(0.04)		(0.02)		(0.04)		(0.02)
High SES		0.429***		0.292**		0.416***		0.288**
		(0.09)		(0.13)		(0.09)		(0.13)
Low SES		−0.106		−0.033		−0.107		−0.055
		(0.08)		(0.12)		(0.08)		(0.12)
Competition choice in S2					0.306***	0.154**	0.398***	0.159*
					(0.08)	(0.07)	(0.09)	(0.08)
Optimism (Relative P.) S2						0.082***		0.083***
						(0.02)		(0.03)
Observations	637	527	665	524	636	524	662	528

Notes: This analysis is based on the control sample only. Coefficient estimates are obtained from ordinary least square estimation where the dependent variable is standardized mathematics grades. Dummies for High SES and low SES are obtained from teacher reported wealth variable. Clustered standard errors (at the school level) are in parentheses. $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

power of the first and second stage competition choice on grades at the baseline. What is noteworthy in this table is that, although the first stage competition choice is not associated with mathematics grades, there appears to be a significant association between second-stage competitiveness and math grades. Even after controlling for cognitive ability, optimism, grit and socioeconomic status, girls (boys) who choose to compete after feedback have about a 0.15 (0.16) standard deviation higher math grades than those who choose not to. Note that grit and optimism are also separately Q5 predictive of math performance, with grit exhibiting considerable predictive power. These results are in line with recent evidence showing that choices in competition tasks predict various labor market and educational outcomes; see Flory et (2014), Reuben, Sapienza, and Zingales (2015), and Buser, Niederle, and Oosterbeek (2014).

What is the reason for the persistent gender gap in the willingness to compete? One explanation is that if girls perform generally worse than boys, we could observe girls (rationally) having a greater tendency to shy away from competition. This would be especially true in the second stage, after receiving performance feedback. However, there is absolutely no gender gap in either piece-rate performance or the first stage performance. As can be seen in Table 4, the performance of girls is in fact slightly higher in all three stages, but the differences do not reach statistical significance. Interestingly, as depicted in Figure 1, the observed gender gap in the second stage seems independent of the type of feedback received in the first stage: girls' willingness to compete in the second stage is lower than boys' among the children that received

TABLE 4. Gender difference in performance (control sample).

	Piece-rate		Stage 1		Stage 2	
	(1) No covariates	(2) With covariates	(3) No covariates	(4) With covariates	(5) No covariates	(6) With covariates
Female	0.238	0.271	0.277	0.295	0.234	0.246
	(0.20)	(0.22)	(0.22)	(0.23)	(0.18)	(0.19)
Male opponent		0.073		0.035		0.193
		(0.15)		(0.22)		(0.21)
Risk tolerance		−0.094		−0.091		−0.048
		(0.08)		(0.09)		(0.06)
Raven score		1.351***		1.351***		1.419***
		(0.11)		(0.16)		(0.15)
Male teacher		−0.388		−0.670		−0.165
		(0.41)		(0.54)		(0.34)
Observations	1353	1174	1356	1178	1357	1178

Notes: This analysis is based on the control sample only. Coefficient estimates are from OLS regressions where the dependent variable is the actual performance (number of correct answers) in the piece-rate stage, stage 1 and stage 2. The first row (female dummy coefficient) gives the estimated performance gender gap. Clustered standard errors (at the school level) are in parentheses. $^{***}p < 0.01$.

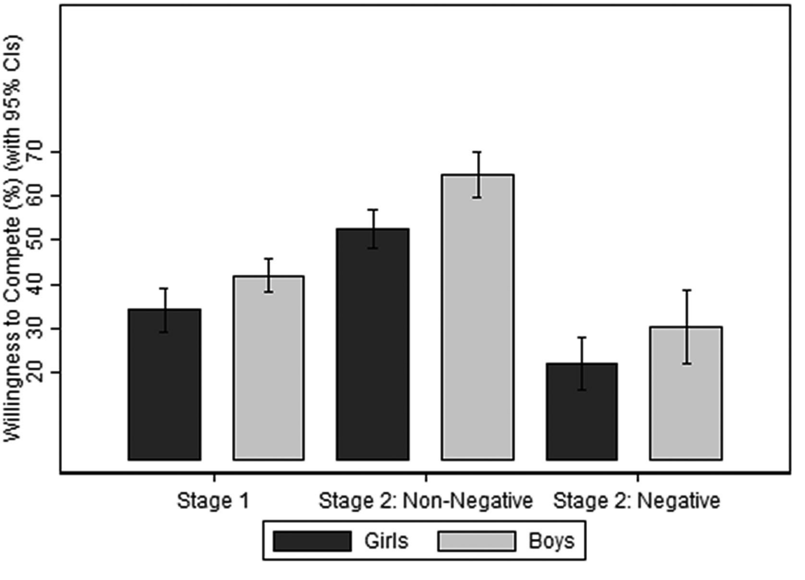


FIGURE 1. Gender difference in the willingness to compete: control sample. The figure presents the proportions of boys and girls who select into the tournament in stage 1 and in stage 2 after non-negative and negative feedback.

positive feedback (by about 10 percentage points) as well as among those that received negative feedback (by about 9 percentage points), with the p-value for the equality of gender gaps across feedback type equal to 0.95.

We now turn to estimating the effect of the treatment on competitive behavior, which is the main goal of the paper. How does the aforementioned educational intervention affect boys' and girls' propensity to compete? Are the gender differences observed in the control group observed in the treatment group?

5.3. Treatment Effects on Competition Choice

In order to test the null hypothesis that the treatment had no impact on the experimental outcome y^s, competition choice in stage $s = \{1, 2\}$, we estimate the average treatment effect by conditioning on baseline covariates:

$$y_{ij}^s = \alpha_0 + \alpha_1 Treatment_j + X_{ij}\gamma + \varepsilon_{ij},$$

where the dependent variable y_{ij}^s is a dummy variable that equals 1 if student i in school j chose to compete in stage s. The binary variable $Treatment_j$ indicates the treatment status of school j, and X_{ij} is a vector of observables for student i in school j that are potentially predictive of the outcome measures we use. These observables include opponent gender (randomly assigned across all students), cognitive score, risk tolerance, grit score collected at baseline, and teacher gender. Note, however, that the estimated effect sizes are not expected to be affected by the inclusion of these covariates because of random assignment, and since they are balanced across treatment groups. The estimated coefficient of the treatment dummy $\hat{\alpha}_1$ yields the average treatment effect on the treated. Estimates are obtained via logit regressions since the outcome considered here is binary. All standard errors are clustered at the level of the school, which is the unit of randomization. In order to account for the fact that we have a small number of clusters (39 schools), we also provide permutation (exact) p-values for the estimated treatment effects. In addition to our two main binary outcomes (first- and second-stage competition choice), in what follows, we look at a number of other outcomes (12 outcomes in total). Since this may raise the issue of multiple hypotheses testing, Table B.2 in the Online Appendix provides Romano–Wolf p-values along with the original ones.

Table 5 presents the estimated treatment effects (average marginal effects) on the willingness to compete for each gender in both the first and the second stage, including the coefficient estimates of the covariates.[23] The first finding to note here is that the first-stage competition choices of all children are very responsive to the treatment (columns (1) and (2)). This shows that willingness to compete is malleable in children. Although the effect size for girls appears to be larger than that of boys, suggesting

23. For the sake of space, in the rest of the tables we will suppress the coefficient estimates of the covariates and present only the estimated treatment effects.

TABLE 5. Treatment effects on competition choice.

	Stage 1		Stage 2	
	(1) Girls	(2) Boys	(3) Girls	(4) Boys
Treatment	0.139***	0.086***	0.152***	0.017
	(0.03)	(0.03)	(0.03)	(0.04)
Male opponent	−0.002	−0.001	0.012	0.001
	(0.02)	(0.03)	(0.02)	(0.03)
Risk tolerance	0.045***	0.039***	0.026**	0.023***
	(0.01)	(0.01)	(0.01)	(0.01)
Raven score	0.050***	0.018	0.049***	0.060***
	(0.01)	(0.02)	(0.02)	(0.01)
Male teacher	−0.016	−0.026	−0.027	−0.053**
	(0.03)	(0.04)	(0.03)	(0.03)
Baseline grit score	0.046***	0.054***	0.072***	0.045**
	(0.01)	(0.01)	(0.02)	(0.02)
Control mean	0.341	0.419	0.399	0.496
Permutation *p*-value	0.001	0.020	0.000	0.701
Girls = boys (*p*-value)	0.130		0.003	
Observations	1256	1312	1255	1313

Notes: Coefficient estimates are average marginal effects from logit regressions where the dependent variable is the binary outcome of competition choice. The average proportions of the control sample (control mean), permutation *p*-values of the estimated treatment effects (permutation *p*-value) and the *p*-value for the equality of treatment effects across gender (girls = boys (*p*-value)) are given at the bottom of the table. Clustered standard errors (at the school level) are in parentheses. $^{**}p < 0.05$; $^{***}p < 0.01$.

some differential treatment effect for girls, the difference does not reach statistical significance (*p*-value for the equality of the effects for girls and boys = 0.13).

Columns (3) and (4) in Table 5 present the estimated treatment effects on the propensity to compete in the second stage. Here, contrary to the first-stage results, we estimate a very strong differential treatment effect with respect to gender (*p*-value for the equality of the effects for girls and boys = 0.003). Specifically, treatment leads to an increase in the propensity to compete in the second stage, after feedback, only for girls. The estimated effect for girls is of considerable size: girls in treated schools are about 15 percentage points more likely to compete in the second stage relative to girls in untreated schools. These results suggest that although the intervention influences the competitiveness of all children, it has differential impact across gender in the second stage, after feedback.[24] We will explore the possible channels through which treatment leads to these gender-differential results in Section 5.5.

Recall that the program we evaluate was implemented first in 2013 and replicated using an independent sample of 16 schools in 2015. Since the two studies we report were independently conducted, the second study gives us the opportunity to assess the

24. Incidentally, we do not estimate statistically significant treatment effects on performance in either stage for either gender. Results are available upon request.

TABLE 6. Treatment effects on competition choice.

	(1) Girls: stage 1	(2) Boys: stage 1	(3) Girls: stage 2	(4) Boys: stage 2
		Sample 1		
Treatment	0.080**	0.005	0.131***	−0.025
	(0.04)	(0.04)	(0.04)	(0.06)
Control mean	0.391	0.442	0.380	0.485
Permutation p-value	0.062	0.918	0.012	0.727
Girls = boys (p-value)	0.130		0.022	
Observations	626	686	629	691
		Sample 2		
Treatment	0.193***	0.154***	0.188***	0.070
	(0.04)	(0.03)	(0.04)	(0.04)
Control mean	0.297	0.395	0.417	0.509
Permutation p-value	0.007	0.003	0.003	0.197
Girls = boys (p-value)	0.389		0.033	
Observations	630	626	626	622

Notes: Coefficient estimates are average marginal effects from logit regressions where the dependent variable is the binary outcome of competition choice. The upper panel presents the results obtained using only sample 1 and the lower panel only sample 2. Covariates used are: opponent gender, risk tolerance, Raven score, teacher gender, and the baseline grit score. The average proportions of the control sample (control mean), permutation p-values of the estimated treatment effects (permutation p-value) and the p-value for the equality of treatment effects across gender (girls = boys (p-value)) are given at the bottom of the table. Clustered standard errors (at the school level) are in parentheses. $^{**}p < 0.05$; $^{***}p < 0.01$.

replicability of our results. Table 6 presents the results separately for each sample. As can be seen in the table, the significant treatment effect we estimate for boys in the first stage is entirely driven by sample 2. We do not estimate a statistically significant treatment effect in either stage for boys in sample 1. However, the effects we document using the pooled sample replicates well for girls: We estimate a statistically significant treatment effect on the first and the second stage competitiveness for girls, with larger estimated effect sizes in sample 2. In fact, we reject the equality of the effect sizes for the first stage competitiveness for both genders across samples (p-value for girls = 0.045, for boys = 0.002). Given that the outcomes were collected four months after implementation of the program for sample 1 and immediately after for sample 2, this difference may imply that the effects weaken over time. However, we find no evidence of such dissipation for the second stage effect on girls (p-value = 0.33). Overall, we interpret these results as inconclusive evidence for the effect of the program on the part of boys and strong evidence for the effect of the program on the part of girls.[25]

The second stage of our experimental task provides us with an outcome measure that is useful for assessing the impact of the intervention on behavior after receiving

25. These results are analogous to studies that consider affirmative action policies to close gender gaps in competition (Balafoutas and Sutter 2012; Niederle, Segal, and Vesterlund 2013), in the sense that girls respond to such policies more strongly than boys as well.

TABLE 7. Treatment effects on stage 2 competition choice by feedback type.

	Non-negative feedback		Negative feedback	
	(1) Girls	(2) Boys	(3) Girls	(4) Boys
Treatment	0.159***	0.027	0.168***	0.049
	(0.03)	(0.04)	(0.04)	(0.05)
Control mean	0.526	0.648	0.220	0.304
Permutation p-value	0.000	0.542	0.003	0.337
Girls = boys (p-value)	0.012		0.023	
Observations	725	693	524	606

Notes: Coefficient estimates are average marginal effects from logit regressions where the dependent variable is the binary outcome of competition choice in the second stage. Covariates used are: opponent gender, risk tolerance, Raven score, teacher gender, and the baseline grit score. Non-negative feedback refers to children who did better than or same as their opponents in stage 1, negative feedback refers to children who did worse than their opponents. Feedback is given to all children regardless of the incentive scheme they chose in stage 1. The average proportions of the control sample (control mean), permutation p-values of the estimated treatment effects (permutation p-value), and the p-value for the equality of treatment effects across gender (girls = boys (p-value)) are given at the bottom of the table. Clustered standard errors (at the school level) are in parentheses. $^{***}p < 0.01$.

performance feedback. Recall that before making the second competition decision, students receive feedback. Specifically, they find out their absolute performance (how many correct answers they had) and their relative performance (whether they did better, worse or equally well relative to their opponent) in the first stage. This feedback is given to everyone, regardless of the choice of incentive scheme in the first stage. Performance in the first competition stage and the feedback received based on this performance are not statistically different across treatment groups (p-value = 0.72 and 0.39, respectively). Based on this finding, we estimate treatment effects on the competition choice in stage 2, conditional on the type of feedback received in stage 1. Table 7 presents the results. Here, we see that the strong treatment effect we estimate for girls (and only for girls) in stage 2 holds true regardless of the type of feedback received in stage 1: the estimated treatment effects across feedback types for girls, which are 15.9 and 16.8 percentage points for non-negative and negative feedback respectively, are statistically no different from each other (p-value = 0.57).

Given the strong differential treatment effects we estimate across gender, the next question is whether the treatment affects the gender gap in competitiveness, in the first and the second stage. Figure 2 depicts the average marginal effects and confidence bands on the female dummy in logit regressions that are run separately for treatment and control, in each of the two stages of competition. Panel 1 depicts unconditional gender gap estimates and panel 2 depicts estimates controlling for all the aforementioned covariates. The main takeaway from these pictures is that evidence for the treatment closing the second stage gender gap is quite strong: the large (9.5 percentage-point) gap estimated in the second stage is completely eliminated in students in treated schools. The evidence on the first stage competition is weaker: while we see that the treatment

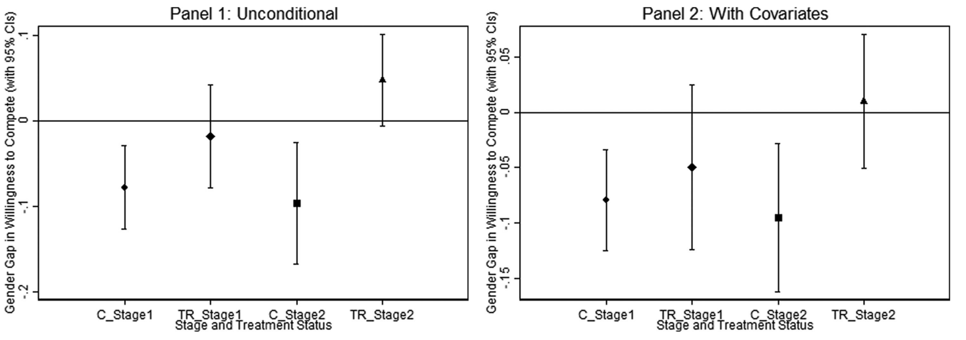


FIGURE 2. Gender difference in the willingness to compete: stage 1 and stage 2. The figure presents the average marginal effects (and confidence bands) on the "female" dummy, obtained via logistic regressions run separately for control and treatment. Panel 1 depicts unconditional gender gaps (raw data) and panel 2 depicts coefficient estimates obtained by controlling for opponent gender, risk attitude, Raven score, and teacher's gender. In both panels confidence bands are obtained via clustered standard errors (at the school level). C: control; TR: treatment.

seems to close the first stage gap as well, the results get weaker once we condition on the covariates.

These results provide evidence that an educational program on fostering grit in the classroom can be quite effective in eliminating both pre- and post-feedback gender gaps in competitiveness. The natural question to ask now is whether the elimination of this gap is necessarily a desired outcome from an economic standpoint, that is, whether it is efficient.

5.4. Efficiency

Based on performance, for some children it is payoff-maximizing to compete, whereas for others it is better to stay out. One can be concerned that interventions such as the one we evaluate in this paper may lead to unintended inferior outcomes for some children, by inducing decisions that turn out to be bad for payoffs ex post. Analyzing how children's decisions fare in terms of expected material payoffs can shed light on these issues. To do this, we first simulate the probabilities of winning and tying in competition for any given performance level, using the empirical distribution of performances. For the first stage, we use the actual performance of the whole sample when calculating the empirical win probabilities. That is, we use simulations to form random matches from the whole empirical distribution in the first choice stage, and compute the empirical win, loss and tie probabilities that correspond to each realized performance level.

For the second stage, however, we utilize the fact that children have received feedback on their first stage performance and will face the same opponent in the next stage. Specifically, this time, we compare children's second-stage performance to the second-stage empirical performance distribution of only better or only worse children

TABLE 8. Treatment effects on payoff-maximizing choices.

	Stage 1		Stage 2		Overall	
	(1) Girls	(2) Boys	(3) Girls	(4) Boys	(5) Girls	(6) Boys
Treatment	0.043	−0.024	0.061***	−0.012	0.084***	0.004
	(0.03)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)
Control mean	0.487	0.561	0.556	0.616	0.304	0.370
Permutation p-value	0.181	0.423	0.025	0.733	0.003	0.899
Girls = boys (p-value)	0.035		0.038		0.006	
Observations	1256	1312	1255	1313	1252	1304

Notes: The coefficient estimates are average marginal effects from logit regressions where the dependent variable is a binary outcome that takes the value 1 if the child made payoff-maximizing choice and zero otherwise. The payoff-maximizing choice indicator is constructed using simulated winning and tying probabilities and actual performances. Covariates used are: opponent gender, risk tolerance, Raven score, teacher gender, and the baseline grit score. The average proportions of the control sample (control mean), permutation p-values of the estimated treatment effects (permutation p-value), and the p-value for the equality of treatment effects across gender (girls = boys (p-value)) are given at the bottom of the table. Clustered standard errors (at the school level) are in parentheses. $^{***}p < 0.01$.

(in the first stage). That is, when calculating the probabilities of winning, rather than using the whole empirical distribution, if a subject was worse (better) than his/her opponent in the first round, we compare his/her observed second stage performance with the second stage performances of only those who were better (worse) than him/her in the first round. Using these probabilities along with realized performances, we calculate each child's expected payoff from competition, and analyze whether the child's actual choice was payoff maximizing ex post. We then estimate the treatment effects on this outcome to see whether the treatment leads to suboptimal choices from a payoff maximization perspective. It should be noted that this analysis does not make utility comparisons and therefore it is not an optimality analysis per se, as it disregards effort costs, which are unobservable.

The first two columns of Table 8 present the estimated treatment effects (average marginal effects) on the propensity to make payoff-maximizing choices in the first stage. Columns (3) and (4) presents second stage results and the last two columns give overall results. The latter is based on the binary variable that takes the value of 1 if the child made a payoff maximizing choice in both stages and zero otherwise. We first note that in the control group, in each stages and overall, girls exhibit lower propensity to make optimal choices from an ex post payoff maximizing perspective: The estimated gender gap overall is about 7 percentage points (in favor of boys), with p-value $= 0.005$. As can be seen from the table, although the evidence for stage 1 is weak, we estimate a statistically significant treatment effect for girls for stage 2 and overall: girls in treated schools are about 8 percentage points more likely than girls in untreated schools to make payoff-maximizing choices in both stages. This analysis provides our first piece of evidence that despite increasing the propensity of tournament entry, the treatment does not cause an inferior outcome for either boys or girls. On the contrary, it improves

TABLE 9. Treatment effects on expected payoff gains.

	Stage 1		Stage 2		Overall	
	(1) Girls	(2) Boys	(3) Girls	(4) Boys	(5) Girls	(6) Boys
Treatment	2.423***	0.978	3.151***	0.011	5.635***	1.033
	(0.74)	(0.69)	(0.88)	(1.17)	(1.37)	(1.58)
Control mean	−1.614	0.444	1.477	4.113	−0.186	4.561
Permutation p-value	0.001	0.183	0.007	0.990	0.000	0.550
Girls = boys (p-value)	0.058		0.016		0.005	
Observations	1250	1304	1243	1289	1241	1281

Notes: The coefficient estimates are coefficients from OLS regressions where the dependent variable is the expected payoff gain. The payoffs are constructed using simulated winning and tying probabilities and actual performances. Covariates used are: opponent gender, risk tolerance, Raven score, teacher gender, and the baseline grit score. The average expected payoff gain for the control sample (control mean), permutation p-values of the estimated treatment effects (permutation p-value) and the p-value for the equality of treatment effects across gender (girls = boys (p-value)) are given at the bottom of the table. Clustered standard errors (at the school level) are in parentheses. $^{***}p < 0.01$.

efficiency by encouraging girls that would do better in competition to compete, without hurting boys. The efficiency improvement generated by the treatment can be further substantiated by looking at the estimated treatment effects on expected payoff gains, as measured by the differences in payoffs between the chosen and the unchosen incentive scheme. This outcome is constructed as the difference between the expected payoff from choosing competition and the (counterfactual) payoff from choosing piece-rate for a child who chooses competition, and vice versa for a child who chooses piece-rate. Note that this measure gives us the child's expected gain from choosing competition over piece-rate (or the other way around), which may be positive or negative ex-post, depending on the performance of the child.[26] Table 9 has the same layout as Table 8. It presents the estimated treatment effects on expected payoff gains for stage 1, stage 2, and overall. We clearly see in this table that girls in treated schools fare very well in terms of expected gains, even in stage 1. Looking at overall results, while the mean expected gain is about −0.186 for the girls in the control group, it is about 5.61 for the girls in the treatment group, which represents a large improvement in efficiency.

Figure 3 summarizes the previous results in terms of gender gaps across treatment status. The first panel depicts the coefficient estimates (average marginal effects) on the female dummy, where the dependent variable is the binary outcome of whether both stage 1 and stage 2 choices were payoff-maximizing or not, separately for treatment and control. Panel 2 depicts the OLS coefficient on the female dummy where the

26. One important point here is that we use the observed performance under the chosen incentive scheme, and therefore abstract from any potential changes in performance in response to (unchosen) incentive schemes (this is also common in the competition literature while calculating ex post payoff costs of choices). However, since performance under competition tends to be at least as high as under piece-rate, our results about girls' under-entry are likely to still stand if performance changes could be taken into account.

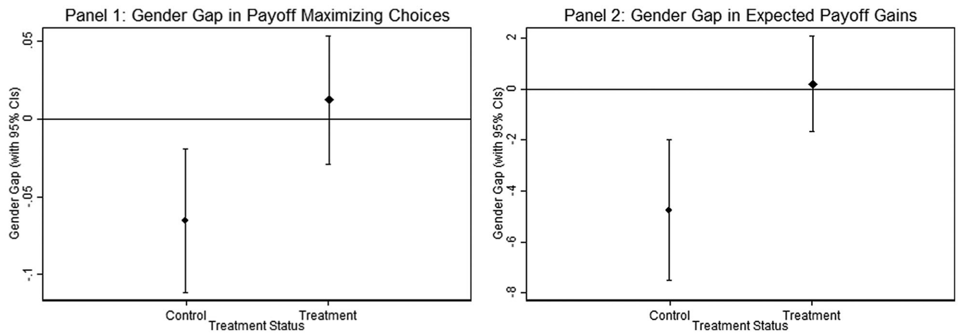


FIGURE 3. Gender difference in payoff-maximizing choices and expected payoff gains. Panel 1 depicts the average marginal effect of the "female" dummy where the dependent variable is the binary outcome of payoff-maximizing choice. Panel 2 depicts the OLS coefficients on the "female" dummy where the dependent variable is expected payoff gains. Confidence bands are obtained via clustered standard errors (at the school level).

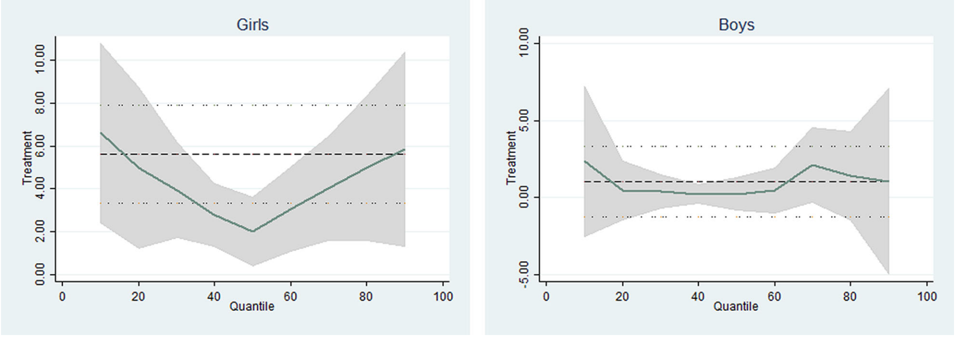


FIGURE 4. Distribution of treatment effects on expected payoff gains. Figures plot the quantile treatment effects on expected payoff gains and 95% confidence bands. The dashed lines indicate the estimated average treatment effects with the dotted lines showing the confidence bands for the average treatment effect estimates.

dependent variable is total expected stage 1 and stage 2 payoff gains. As clearly seen from the figure, although there is a significant gender gap in the probability of making payoff maximizing choices in the control group, this gap is completely erased by the treatment (panel 1). Similarly, the gap in favor of boys with respect to expected payoff gains is eliminated by the treatment (panel 2).

Although these results are quite positive, we acknowledge that these are average effects, which may conceal undesired distributional effects. In order to see the distributional impact of the treatment, we estimate quantile treatment effects on expected overall payoff gains. The two panels in Figure 4 depict the coefficient estimates and 95% confidence bands for expected payoff gains for boys and girls. At first glance, the figures show that treatment effects are generally non-negative across quantiles, for both genders, reassuring us that no one was hurt by the treatment in terms of payoffs. For girls, the effects are positive across quantiles, and statistically significant in the low and high quantiles. This is because very low gains (large negatives) result from under-entry into the tournament by girls that would have done well in the tournament, which is significantly lessened by the treatment. Positive treatment effects at high gains are indicative of optimal entry into competition (from a payoff-maximization perspective), which is significantly more prevalent in the treatment group. Taken together, these results suggest that the increase in competitiveness brought about by the intervention does not harm girls (or boys) in terms of payoffs at any point in the distribution. On the contrary, the intervention results in significant payoff gains for girls, suggesting that many able girls are staying away from competition in the baseline.

Further support for positive distributional effects comes from our analysis of heterogeneous treatment effects based on cognitive ability, as measured by Raven's progressive matrices. Specifically, we re-estimate the treatment effects on stage 1 and stage 2 choices for students with below and above median cognitive score. Table B.3 in the Online Appendix presents the results of this analysis for males and females separately. The estimates presented in this table clearly show that the treatment effects on competition choices are similar for low and high cognitive scorers within females and within males. A notable finding here is that the treatment seems to be very effective in inducing both groups of girls to choose to compete, in both stages. This immediately brings to mind that girls with low cognitive ability in treated schools may fare worse than those in the control group in terms of expected gains. However, Table B.4 in the Online Appendix shows that this concern is unwarranted. We observe that girls with high cognitive ability in treated schools are significantly more likely to make payoff-maximizing choices in both stages than high ability girls in control (9.1 percentage points), and this holds true for low ability girls as well (an 8 percentage point effect). Note also that we estimate a positive and significant treatment effect on expected payoff gains for both groups of girls (see columns (5) and (6)).

Overall, these results suggest that the treatment corrects a significant inefficiency by inducing girls, who were under-entering the tournament in the baseline, to compete. The fact that we do not estimate significant treatment effects on payoff-maximizing choices or payoff gains for boys likely suggests that boys' pretreatment decisions were generally optimal from a payoff maximizing perspective and the treatment did not have an adverse effect (any effect) on their decisions and payoffs. In the next section, we will explore potential mechanisms that may be responsible for these gender-differential effects of the treatment.

5.5. Mechanisms of Impact

Which aspects of the program are likely to drive girls into competition and eliminate the gender gaps observed in both stages in the baseline? One major potential channel that may have led to the effects we estimate for girls may be altered expectations about future performance. These expectations/beliefs capture how optimistic the individual is about how well he/she can do in the next stage. Our design allows us to explore this channel using a number of useful measures within the task. Our first measure (already described in Section 5.2) is children's elicited expected performance relative to her opponent, which we refer to as optimism about relative performance. Recall that we have this measure for both stages. Our second measure, which is expected to capture optimism about how much one can improve on a past performance, is constructed as the difference between the child's expected second stage performance and her actual first stage performance. We refer to this measure as optimism about performance improvement. The idea behind using this measure is that optimism about improvements in own performance captures an important component of the treatment material, which is the optimistic (and individualistic) view that one can always do better than before with sustained effort. Naturally, we have this measure only for the second stage. In addition to expectations/beliefs, we explore a likely behavioral channel, which is based on the conjecture that the treatment induces competitive behavior through its effect on children's grit/perseverance. In the first stage, a gritty child is likely to set an ambitious performance goal, and in the second stage, a gritty child who believes in the return to perseverance may choose to compete even after a failed attempt.

In what follows, we will consider a number of causal mediation models where each of the aforementioned variables (and possibly, their interactions) can be a potential mediator. In doing this, we will restrict our attention to only girls. This is because we have robust treatment effects on girls in both stages in both samples, which make it meaningful to explore mediation channels. In contrast, there is no treatment effect on boys' competitiveness in the second stage in either sample, and although we estimate a significant treatment effect for boys in the first stage in sample 2, this does not replicate in sample 1, rendering the evidence on boys inconclusive. We provide full analyses, including boys in the Online Appendix. In addition to the effects we estimate on competitiveness, several useful facts regarding our candidate mediators also give support to our choice of focusing on girls. Table 10 presents estimated treatment effects on optimism about relative performance, optimism about performance improvement and the post-treatment grit score. Here, we see that the effect sizes of all these measures are stronger for girls, although we fail to reject equality across gender, except for the grit score.[27]

We conduct causal mediation analyses using two IV-based techniques. The first one is a widely used standard IV estimation where the causal effect of an endogenous mediator on a final outcome is estimated. In our case, this involves estimating the causal effect of say, grit on the willingness to compete by instrumenting the post-treatment grit score by the treatment indicator. Alternatively, another model that assumes a different endogenous mediator, say optimism, can be estimated in the same way. Given that

27. Incidentally, conditional on actual performance, girls have more pessimistic beliefs about their future performance relative to their opponents at the baseline. The gap in optimism about future relative performance is about 0.12 standard deviations (p-value = 0.046) in the first stage and is about 0.10 standard deviations (p-value = 0.044). The treatment closes these gaps. Using our pretreatment grit scores, we observe that girls report significantly more gritty behaviors than boys (in both treatment and control). However, controlling for the pretreatment grit score, although we do not see a gender difference in the post-treatment grit score in the control group, we estimate a significant increase in grit on the part of girls due to treatment, generating a significant gender gap in favor of girls.

TABLE 10. Treatment effects on potential mediators.

	Optimism (Relative P.: S1)		Optimism (Relative P.: S2)		Optimism (Imp. P: S2)		Grit score	
	(1) Girls	(2) Boys	(3) Girls	(4) Boys	(5) Girls	(6) Boys	(7) Girls	(8) Boys
Treatment	0.161**	0.131	0.122**	0.062	0.160**	0.128	0.387***	0.249***
	(0.07)	(0.08)	(0.06)	(0.07)	(0.08)	(0.09)	(0.06)	(0.07)
Control mean	−0.123	−0.015	−0.077	−0.009	−0.105	−0.047	−0.044	−0.221
Permutation p-value	0.045	0.138	0.063	0.366	0.071	0.204	0.000	0.004
Girls = boys (p-value)	0.745		0.389		0.700		0.052	
Observations	1247	1284	1242	1285	1246	1288	1088	1117

Notes: The coefficient estimates in columns (1) and (2) are coefficients from OLS regressions where the dependent variable is expected relative performance in stage 1, the estimates in columns (3) and (4) are coefficients from OLS regressions where the dependent variable is expected relative performance in stage 2, the estimates in columns (5) and (6) are coefficients from OLS regressions where the dependent variable is the difference between expected second stage performance and actual first stage performance. The coefficient estimates in the last two columns are from OLS regressions where the dependent variable is the post-treatment grit score. Covariates used are: opponent gender, risk tolerance, Raven score, teacher gender, and the baseline grit score. Average outcomes for the control sample (control mean), permutation p-values of the estimated treatment effects (permutation p-value), and the p-value for the equality of treatment effects across gender (girls = boys (p-value)) are given at the bottom of the table. Clustered standard errors (at the school level) are in parentheses. $^{**}p < 0.05$; $^{***}p < 0.01$.

TABLE 11. Mediation analysis: single mediator IV.

	Stage 1		Stage 2		
	(1) Comp: S1	(2) Comp: S1	(3) Comp: S2	(4) Comp: S2	(5) Comp: S2
Optimism (Relative P.) S1	0.305*** (0.01)				
Grit score		0.239*** (0.02)			0.247*** (0.02)
Optimism (Relative P.) S2			0.323*** (0.01)		
Optimism (Improved P.) S2				0.302*** (0.02)	
Observations	1243	1086	1238	1242	1084

Notes: The analysis is based on girls only. The coefficient estimates are from IV probit estimation where the dependent variables are stage 1 and stage 2 competition choice (girls), respectively. Column (1): optimism (expected relative performance stage 1) is instrumented with the treatment indicator (dependent variable is competition choice in stage 1). Column (2): post-treatment grit score is instrumented with the treatment indicator (dependent variable is competition choice in stage 1). Column (3): optimism (expected relative performance stage 2) is instrumented with the treatment indicator (dependent variable is competition choice in stage 2). Column (4): optimism (expected performance improvement) is instrumented with the treatment indicator (dependent variable is competition choice in stage 2). Column (5): Post-treatment grit score is instrumented with the treatment indicator (dependent variable is competition choice in stage 2). Covariates used are: opponent gender, risk tolerance, Raven score, teacher gender, and the baseline grit score. Clustered standard errors (at the school level) are in parentheses. $^{***}p < 0.01$.

we have only one instrument (treatment assignment), this approach imposes a strong exclusion restriction and assumes that (i) the proposed mediator is the sole (correct) mediator of the treatment effect, and (ii) the treatment does not have a direct effect on the final outcome. Table 11 presents the results for the three mediators: optimism (relative performance) for stage 1 and stage 2 , optimism (improved performance, for stage 2) and the post-treatment grit score, for both stage 1 and stage 2. As we see in the table, all four mediators have very strong effects on girls' competitiveness, in both the initial choice stage (when relevant) and after feedback. Take stage 2 competition for example. A one standard deviation increase in optimism regarding performance improvement leads to about 30 percentage points increase in the probability of competing.

Although the previous single IV mediation model provides evidence, albeit suggestive, on the causal effects of our proposed mediators on competitiveness, a mechanism where the effect of a mediator runs through another intermediate mediator may be more appropriate in our case. What we have in mind here is that, in addition to independently mediating the treatment effect by encouraging a perseverant attitude toward challenging tasks, increased grit may have led one to expect a higher performance from herself, which in turn results in higher willingness to compete. Such a mechanism requires identifying the causal effect of the mediator (grit) on the final outcome (competitiveness), mediated by an intermediate mediator (such as optimism about performance improvement), as well as its direct effect, using a single instrument. Dippel et al. (2017) propose a framework to estimate such a mechanism. The framework involves first estimating the effect of the endogenous mediator on the final outcome (the total effect) and decomposing this total effect into an "indirect effect" running though the intermediate mediator and a "direct effect". Note that this model also imposes a very strong exclusion restriction by assuming that grit is the only mediator and the treatment does not have a direct effect on the final outcome. The advantage of this method over the standard IV method, however, is that with a single instrument, the causal effect of an intermediate mediator on the final outcome can be also estimated.

Denote our final outcome, competitiveness as Y, our endogenous mediator (post-treatment grit) as G and our intermediate mediator (optimism) as O. The first step is to estimate the effect of the mediator on the final outcome using an IV estimator where the mediator is instrumented with the treatment indicator; denote the second-stage coefficient as θ_G^Y, which gives the "total effect" of the mediator on competitiveness. The second step is to estimate (again, using an IV estimator) the effect of the mediator on the intermediate mediator; denote the second stage coefficient as θ_G^O. The final step involves another IV estimation where the effect of the intermediate mediator on competitiveness is estimated conditioned on the mediator where the second stage coefficient on the mediator (denoted as $\theta_G^{Y|G}$) gives the "direct effect" of the mediator and the second stage coefficient on the intermediate mediator ($\theta_O^{Y|G}$) multiplied with θ_G^O gives the "indirect effect" of the mediator through its effect on the intermediate mediator. Notice that the final step relies on a strong assumption that conditional on the endogenous mediator, the treatment is independent of potential outcomes; see full details and proofs in Dippel et al. (2017).

The two panels in Table 12 present the results for stage 1 and stage 2. In both panels the mediator is the post-treatment grit score. The first panel aims to explain the effect on the initial choice of competition, and assumes that the intermediate mediator is optimism (about relative performance) in the first stage. The results show that this proposed causal mechanism yields a statistically significant indirect effect of grit on the first stage competition choice. We estimate a statistically insignificant direct effect of grit for this stage. Put quantitatively, we estimate an average causal mediated effect (indirect effect) of about 37.7 percentage points, which corresponds to 95% of the total 39.8 percentage point effect on first-stage competitiveness.

The second panel focuses on the post-feedback choice of competition, and assumes that the intermediate mediator is optimism regarding one's own performance improvement.[28] Here, we estimate an indirect effect of about 33.8 percentage points (about 79%) of the total 42.8 percentage point effect on competitiveness. Put more specifically, a one standard deviation increase in grit induces about a 0.41 standard deviation increase in optimism about performance improvement, which in turn

28. We did consider expected relative performance in stage 2 as an intermediate mediator in the second stage as well; however, this mediation model yields insignificant indirect effect, likely due to low variation in this variable. Results for all considered mediation models are given in the Online Appendix.

TABLE 12. Mediation analysis: intermediate mediator IV.

	Stage 1			Stage 2		
	(1) G→Y	(2) G→O	(3) G→O→Y	(4) G→Y	(5) G→O	(6) G→O→Y
Grit score	0.398***	0.509**	0.009	0.428***	0.407*	0.098*
	(0.10)	(0.23)	(0.02)	(0.11)	(0.24)	(0.05)
Optimism (Relative P.) S1			0.740***			
			(0.28)			
Optimism (Improved P.) S2						0.830*
						(0.43)
Indirect effect			0.377**			0.338**
SE (indirect effect)			(0.100)			(0.108)
Observations	1086	1080	1078	1084	1078	1075

Notes: The analysis is based on girls only. The coefficient estimates in columns (1) and (4) total effect of grit in stage 1 and 2, respectively. They are obtained from two stage least square estimation where the dependent variables are stage 1 and stage 2 competition choice (girls), respectively. The estimates in columns (2) and (5) are obtained via two stage least square regressions of optimism (relative performance) and optimism (own performance), respectively, on post-treatment grit score where the latter is instrumented with the binary treatment indicator. Finally, columns (3) and (6) give the coefficients from two stage least square regressions where optimism about relative performance (stage 1) and optimism about own performance (stage 2) are instrumented with the treatment indicator and the post-treatment grit score is added to the list of exogenous variables. The product of the coefficient in column (2) (5) and the coefficient on optimism in column (3) (6) gives the average causal mediated effect (indirect effect) and these are provided at the bottom of the table along with their standard errors in parentheses. Clustered standard errors (at the school level) are in parentheses. $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

increases the propensity to compete in the second stage by 33.8 percentage points (0.407 times 0.83). Notice that we also estimate a statistically significant direct effect of grit (about 10 percentage points) for this stage.[29]

6. Concluding Remarks

Documenting, understanding the reasons for, and exploring policies to mitigate gender gaps in competitiveness has been a very active area of research in economics in recent years. Considering competitive behavior in a dynamic context where individuals choose an incentive scheme, receive performance feedback and make a decision again, we document that grit is one of the driving forces of sustained willingness to compete.

29. It is important to re-emphasize that one can conduct other types of mediation analyses, entertain the possibility of other causal mechanisms and assume different mediators (and intermediate mediators). One such potential mediator, for example, is risk-tolerance. Although we do not have these data for sample 2, we are able to compare risk tolerance across treatment and control groups for sample 1. In unreported regressions, we find that the treatment does not have a significant effect on risk tolerance ($p = 0.36$). This suggests that the impact of the intervention on competitive behavior does not work through increasing the tolerance for risk. Still, we acknowledge that there may be other potential channels that our data do not allow us to consider.

We then explore whether competitive behavior can be influenced by an educational intervention that aims to foster grit in the classroom environment via a carefully designed curriculum that is implemented by trained teachers. Using choices in a two-stage competition task as our outcome variable, we evaluate the impact of this unique educational intervention.

We first show that boys and girls have a significantly different propensity to compete, even in childhood. We then show that an educational intervention that aims to foster the noncognitive skill of grit can have a significant impact on both girls' and boys' competitiveness. This indicates that competitiveness as an individual attitude is malleable in childhood through a targeted program that promotes ambitious goal-setting and perseverance. The effect on competitiveness naturally has implications for the gender gap as well. We estimate a statistically significant treatment effect on the willingness to compete for both boys and girls in the first stage and this effect is bigger for girls, weakly mitigating the first stage gender gap. Moreover, we find a significant treatment effect on competitiveness after performance feedback only for girls, entirely eliminating the second stage gender gap. Our data suggest that improved grit and optimism on the part of girls is likely to be the main mechanism that generates our results.

Coupled with the fact that the treatment is overall payoff-improving for girls and does not hurt boys, our results indicate that fostering grit in the classroom environment can address one of the most debated gender gaps in the literature without sacrificing efficiency. Incorporating the type of intervention proposed in this paper into the classroom environment (or even the home environment) can be an easy and cost-effective alternative from a policy perspective. In addition to short-run effects, such an intervention may potentially influence important real outcomes such as track choices in education, if effects persist over the years. Studying these longitudinal effects are left for further research. Anecdotal evidence based on testimonies from participating teachers suggests that teachers' own beliefs, attitudes and classroom practices such as feedback styles (e.g., praising effort) have likely changed in response to the treatment. Although our phase-in design does not make it possible in the current sample, another interesting avenue for further research would be studying the effect of such changes in teaching practices and philosophy, using new cohorts of students that do not directly receive the educational materials, but are taught by previously treated teachers.

Finally, our data highlight the relevance of studying competitiveness in a dynamic framework where, as in real life, individuals receive performance feedback and decide whether or not to stick with their chosen path. Performance feedback is ubiquitous in economic life and in education, and the way it is interpreted may have profound impact on an individual's subsequent decisions. Evaluations of educational and other policies that aim to mitigate inefficient gender gaps in competitiveness in particular, and achievement in general, will be more comprehensive if their effects on dynamic choices and their interactions with feedback are taken into account.

Appendix

Q6

Questions Used for Constructing the Grit Score

Four-point item scale: completely agree, agree, disagree, completely disagree

(1) I like schoolwork best which makes me think hard, even if I make a lot of mistakes.
(2) Setbacks discourage me.
(3) If I think I will lose in a game, I do not want to continue playing.
(4) If I set a goal and see that it's harder than I thought I easily lose interest.
(5) When I receive a bad result on a test I spend less time on this subject and focus on other subjects that I'm actually good at.
(6) I work hard in tasks.
(7) I prefer easy homework where I can easily answer all questions correctly.
(8) If I'm having difficulty in a task, it is a waste of time to keep trying. I move on to things which I am better at doing.

Instructions for the Competition Task

Today we are going to play a game with you. At the end of the game, you will get to choose gifts from this gift bag [show gift bag]. You will be asked to make some decisions in the game. How many gifts you earn will depend on your decisions, and your performance in a task. We will now explain the rules for this. Please listen very carefully. You can raise your hand if you have questions about the rules. But in all the decisions you make, you should think to yourself and not tell your decision out loud to others, OK? There is no right or wrong decision in this game. Everyone is free to choose what he/she wants.

The game will consist of three parts: part 1, part 2, and part 3. Only one of these parts will count to determine your rewards. At the very end of the game, we will have a draw. Here are pieces of paper with 1, 2, and 3 written on it. [Show papers] We will put these in a bag and draw one. Whichever part comes out, only your decisions and performance in that part will count for rewards. This means, the gifts do not accumulate in every part. You get rewards only based on what happens in the randomly drawn part. Let's give an example. Suppose someone would earn 3 gifts from the 1st part, 5 from the 2nd and 4 from the 3rd. In the draw at the end, part 2 came out. How many gifts does this person actually get? 5. Not 12, because the gifts do not accumulate. Any questions? OK.

Now, the task in this game will be to add numbers. In each question, there are two 2-digit numbers and 1 one-digit number to be added. For example, [written on the board] 75+80+4=?, 72+47+8=?. You will be given many addition questions like this. Now, we will distribute workbooks. There are many pages in this workbook. It is extremely important that you do not turn over a page before we tell you to do so, or

TABLE A. Evaluation Design.

	STUDY 1 Patience+Grit (9 schools)	Patience (6 schools)	Control (8 schools)	STUDY 2 Grit (8 schools)	Control (8 schools)
Baseline Data	March 2013	March 2013	March 2013	May 2015	May 2015
Patience Treatment	Spring 2013	Fall 2013	-	-	-
Grit Treatment	Fall 2013	-	-	Fall 2015	-
Follow-up Data	May 2014	May 2014	May 2014	Jan 2016	Jan 2016

go back to a previous page. Everyone should be at the same page at any given time. OK? We are also going to give you a separate sheet of paper with some questions on it. First, just write your name on everything you get. Then put the single sheet of paper aside with written side down. That is for later. Once you write your name, please wait and do not open any pages. Now, I will explain the rules for the 1st part of the game. In this part, you will be given 2.5 min. During this time, you will try to do as many addition questions as you can. There will be many such questions on your sheet. If this part (part 1) gets drawn at the end of the game, you will earn gifts as follows: every question you solve correctly in this part will mean one gift of your choice from the gift bag. The more questions you solve, the more gifts you get. If you do 1 right, you get 1 gift, if you do 4 you get 4, if you do 7, you get 7, so on and so forth. Is this clear? [If no questions] Are you ready? OK, on the count of three, everyone will turn the page to start. 3, 2, 1, start! [At the end of 2.5 min] OK, everyone turn the page! You should now see the waiting page, page 2. Stay there, OK?

Now, we played this same game, before you, in other classes in other schools. Those children were also 4th graders like you are. We also timed them on the exact same task with 2.5 min, and we recorded how many sums each student did correctly in all three parts of the game. Now, we will ask you something about the task you just did. Remember, you will think to yourself and not say anything out loud, OK? We randomly selected, from those other classes in other schools where we played the game before, one boy and one girl. One is called Batuhan, another called Merve [note: Batuhan is a common boys' name, Merve a common girls' name in Turkey]. Now, we will ask you to guess: which of them do you think did more questions right in 2.5 min? Batuhan more, Merve more, or did they do equally well? If you guess correctly, you will get an extra, smaller gift. OK? Now turn the page. You'll be on page 3. Circle your guess there, and turn the page [children see a waiting page] Now, remember we told you we did this game in other classes in other schools before. We will match each one of you, randomly, with one student from that group. You will not know who that student is. You will never see them, you will not get to meet them. For example, let's say, one of you, got matched with somebody from another class, called Hakan. We had given Hakan a number when we visited that class. Say, student number 1024. You will not know that you got matched with Hakan. The only thing you will know is whether your matched student is a boy or a girl. Let's take another one of you. Suppose you were matched with student number 1038. You will not know who this person is. But you will know whether it's a girl or a boy. OK? Keep this in mind for now.

Now, we will explain the rules of the 2nd part of the game. In this part, the task is the same. You will again do addition questions. And again, you will have 2.5 min. However, in this part, you will choose how you earn gifts. You have two options: One option is "per-question". [Write on board: per-question] What does this mean? If you choose this, for every addition question you answer correctly you earn one gift. That is, 2 questions right = 2 gifts, 4 questions right = 4 gifts, 7 questions right = 7 gifts, so on and so forth. [Write on board: Every correct answer = 1 gift] The other option is "competition". [Write on board: Competition] Remember that every one of you is going to be matched with another student? If you choose competition, you will be competing with that matched student. In this case, how many gifts you earn will depend on how many questions you do correctly and how many the matched student did correctly. We have here, recorded, the number of correct questions solved by each of the matched students in each part of the game. If you do more questions right than your matched student in part 2, that is, if you outperform that student, you get not 1, but 3 gifts for each question you solve correctly. [Write on board: Better: Each correct answer = 3 gifts] But if you do less questions right than your matched student, that is if he/she did better than you, you get zero gifts [Write on board] If you do an equal number of correct questions as your matched student, you get 1 gift for each question you solve correctly. [Write on board: Tied: Each correct answer = 1 gift] Is this clear? Now, let's give some examples. Suppose that I chose the per-question option. Suppose that I had 5 questions correct. [Write performances on board] How many gifts do I get? $5 \times 1 = 5$. It does not matter what my matched student had done. Suppose that I chose to compete. How many gifts do I get? Suppose my matched student had 4 questions correct. I outperformed the matched student, so I get $5 \times 3 = 15$ gifts. What if the matched student had 7 questions correct? The matched person outperformed me, so I get 0 gifts. What if the matched student had 5 questions correct, like I did? It is a tie, so I get $5 \times 1 = 5$ gifts. [Give 2 more examples, make sure students fully understand] Now let's give another example. Suppose that I chose piece-rate and got 7 questions right. How many gifts do I get? $7 \times 1 = 7$ gifts. It is 7 regardless of what my matched student's performance. Now suppose that I chose competition. Suppose my matched student had less than 7 questions right. How many gifts do I get? My performance was better, so I get $7 \times 3 = 21$ gifts. Suppose that my matched student had more questions right than me. What do I get? I get 0 gifts. Suppose that my matched student also had 7 questions right. What do I get? $7 \times 1 = 7$ gifts. Did everyone understand? Do you have any questions?

Now, on your workbooks, there will be an ID number corresponding to the student you were matched with. And you will see if this person is a boy or a girl. Now everyone turn the page. Do you see it? Now please think carefully and silently. If part 2 is chosen for payment, based on which option would you like to get gifts? Please circle what you want, and turn the page. OK. Now, we will ask you two questions. As you know, every one of you is matched with a student from another school and class. Everyone, even if they did not choose to compete, has a matched student. Now, please think. Every part of this game has 16 addition questions. How many questions do you think you will get right in part 2? And how many do you think your matched student got right? Think carefully because every correct guess can get you an extra small gift. At the end of the game, we will pick 3 students among you randomly. We will look at their guesses. For each correct guess, we will give them an extra small gift. Understood? Now please everyone turn the page and make your guesses. Is everyone done? Now, you will start doing the questions in part 2 and will have 2.5 min again. One thing to note–you should not stop solving questions just because you've already reached your guess and want to be consistent with it. Why? Because solving more questions brings you more gifts than being consistent with your guess. So to get more gifts, you should try to do as many as you can in the time you are given. OK? Is everyone ready? Turn the page, and start on the count of 3. [When time is up] Everyone turn the page immediately. You'll see page 7 (a waiting page).

Now, we will come to each of your desks. We will calculate how many questions you did, and compare it with your matched student's performance. We will mark on your sheet how many you did correctly. We will also mark whether you have done more, the matched student has done more, or you both did the same number of correct questions. Everyone will get this information. You should be extremely quiet here and not tell anyone the information you got. Now while you are waiting, please take that separate sheet we distributed at the start. There are some questions for you there. Please start filling that in. [Note: Students work on a short questionnaire to fill the time when they are waiting for their performance to be calculated and feedback given. When marking, empty answer blanks were crossed over with a pen so that the child cannot go back and change performance. Experimenters wrote the number of correctly solved questions and used a code on the workbooks to make the performance comparison, and circled the outcome of the comparison on the student's feedback page.] So, did everyone see how many they got right? Did everyone see whether they did more, less or equal number of questions with their matched student? You know how many you got right. You know whether it was more, less or equal. But notice, you do not know how many more or how many less you did than the matched student.

Now, we will start the 3rd part of the game. In this part, you will again do addition problems in 2.5 min. Your matched student is the same student as in part 2 for this part. Remember that those matched students also did this same task for 3 periods. We also recorded how many each matched student did correctly in the 3rd part. Now please think. Again, you have 2 options for how you earn gifts. You can choose per question, or you can choose to compete with the matched student. If you choose per-question, for every correct answer, you get 1 gift. If you choose to compete, you get 3 gifts per correct answer you have, but only if you do more questions than your matched student did in part 3. If you do less, you get zero gifts. If you do equal, you get 1 gift per correct answer. Now, everybody turn the next page. There you will make your decision for part 3. If part 3 is drawn at the end of the game for reward, you will get gifts according to what you choose here. Did everyone decide? OK, now turn the page. You will again make guesses. How many questions do you think you will get right in part 3? How many questions do you think your matched student did in part 3? Don't forget, again, every correct guess can get you an extra small gift if you are one of the 3 people chosen. Now make your guesses and wait, do not turn the page. Everyone ready? OK, now on the count of 3, turn the page, and start! [When time is up] Everyone stop and turn the page!

References

Alan, S. and S. Ertac (forthcoming). "Fostering Patience in the Classroom: Results from a Randomized Educational Intervention." *Journal of Political Economy*. Q7

Alan, S., T. Boneva, and S. Ertac (2017). "Ever Failed, Try Again, Succeed Better: Results from a Randomized Educational Intervention on Grit." HCEO Working paper. Q8

Almas, I., A. Cappelen, K. G. Salvanes, E. O. Srensen, and B. Tungodden (2016). "Willingness to Compete: Family Matters." *Management Science,* 62, 2149–2162.

Almlund, M., A. L. Duckworth, J. J. Heckman, and T. D. Kautz (2011). "Personality Psychology and Economics." *In Handbook of the Economics of Education* edited by E. Hanushek, S. Machin, and L. Woessman. Elsevier, Amsterdam, pp. , 1–181. Q9

Andersen, S., S. Ertac, U. Gneezy, J. A. List, and S. Maximiano (2013). "Gender, Competitiveness, and Socialization at a Young Age: Evidence from a Matrilineal and a Patriarchal Society." *Review of Economics and Statistics*, 95, 1438–1443.

Andersen, S., S. Ertac, U. Gneezy, A. J. List, and S. Maximiano (2014). "Gender, Competitiveness, and the Response to Performance Feedback: Evidence from a Matrilineal and a Patriarchal Society." Working paper.

Azmat, G. and N. Iriberri (2010). "The Importance of Relative Performance Feedback Information: Evidence from a Natural Experiment using High School Students." *Journal of Public Economics*, 94, 435–452.

Balafoutas, L. and M. Sutter (2012). "Affirmative Action Policies Promote Women and Do Not Harm Efficiency in the Laboratory." *Science*, 335, 579–582.

Barankay, I. (2011). "Rankings and Social Tournaments: Evidence from a Crowd-sourcing Experiment." Wharton School of Business, University of Pennsylvania.

Berlin, N. and M. P. Dargnies (2016). "Gender Differences in Reactions to Feedback and Willingness to Compete." *Journal of Economic Behavior and Organization*, 130, 320–336.

Bertrand, M. (2011). "New Perspectives on Gender." *Handbook of Labor Economics*, 4, 1543–1590.

Bertrand, M. and K. F. Hallock (2001). "The Gender Gap in Top Corporate Jobs." *Industrial and Labor Relations Review*, 55, 3–21.

Blau, F. D. and L. M. Kahn (2000). "Gender Differences in Pay." *Journal of Economic Perspectives*, 14, 75–99. Q10

Blau, F. D., M. A. Ferber, and A. E. Winkler (2002). *The Economics of Women, Men and Work*. Prentice Hall, Englewood Cliffs, NJ.

Booth, A. and P. Nolen (2012). "Choosing to Compete: How Different are Girls and Boys?" *Journal of Economic Behavior and Organization*, 81, 542–555.

Borghans, L., A. L. Duckworth, J. J. Heckman, and B. ter Weel (2008). "The Economics and Psychology of Personality Traits." *Journal of Human Resources*, 43, 972–1059.

Buser, T., M. Niederle, and H. Oosterbeek (2014). "Gender, Competitiveness and Career Choices." *Quarterly Journal of Economics*, 129, 1409–1447.

Cárdenas, J. C., A. Dreber, E. Von Essen, and E. Ranehill (2012). "Gender Differences in Competitiveness and Risk Taking: Comparing Children in Colombia and Sweden." *Journal of Economic Behavior and Organization*, 83, 11–23.

Croson, R. and U. Gneezy (2009). "Gender Differences in Preferences." *Journal of Economic Literature*, 448–474.

Dippel, C., R. Gold, S. Heblich, and R. Pinto (2017). "Instrumental Variables and Causal Mechanisms: Unpacking the Effect of Trade on Workers and Voters." NBER Working Paper No. 23209.

Dreber, A., E. Von Essen, and E. Ranehill (2014). "Gender and Competition in Adolescence: Tasks Matter." *Experimental Economics*, 17, 154–172.

Duckworth, A. L., C. Peterson, M. D. Matthews, and D. R. Kelly (2007). "Grit: Perseverance and Passion for Long-Term Goals." *Journal of Personality and Social Psychology*, 92, 1087–1101.

Duckworth, A. L. and P. D. Quinn (2009). "Development and Validation of the Short Grit Scale (Grit-S)." *Journal of Personality Assessment*, 91, 166–174.

Duckworth, A. L., T. A. Kirby, E. Tsukayama, H. Berstein, and K. A. Ericsson (2011). "Deliberate Practice Spells Success: Why Grittier Competitors Triumph at the National Spelling Bee." *Social Psychology and Personality Science*, 2, 174–181.

Eriksson, T., A. Poulsen, and M. C. Villeval (2009). "Feedback and Incentives: Experimental Evidence." *Labour Economics*, 16, 679–688.

Ertac, S. and B. Szentes (2011). "The Effect of Information on Gender Differences in Competitiveness: Experimental Evidence." TÜSİAD-Koç University Economic Research Forum Working paper series.

Eskreis-Winkler, L., E. P. Shulman, S. Beal, and A. L. Duckworth (2014). "Survivor Mission: Why Those Who Survive Have a Drive to Thrive at Work." *Journal of Positive Psychology*, 9, 209–218.

Flory, J. A., A. Leibbrandt, and J. A. List (2015). "Do Competitive Workplaces Deter Female Workers? A Large-Scale Natural Field Experiment on Job-Entry Decisions." *The Review of Economic Studies*, 82, 122–155.

Gill, D. and V. Prowse (2014). "Gender Differences and Dynamics in Competition: The Role of Luck." *Quantitative Economics*, 5, 351–376.

Gill, D., Z. Kissová, J. Lee, and V. L. Prowse. (forthcoming). "First-Place Loving and Last-Place Loathing: How Rank in the Distribution of Performance Affects Effort Provision." *Management Science*.

Gneezy, U., K. L. Leonard, and J. A. List (2009). "Gender Differences in Competition: Evidence from a Matrilineal and a Patriarchal Society." *Econometrica*, 77, 1637–1664.

Gneezy, U., M. Niederle, and A. Rustichini (2003). "Performance in Competitive Environments: Gender Differences." *Quarterly Journal of Economics*, 118, 1049–1074.

Gneezy, U. and J. Potters (1997). "An Experiment on Risk Taking and Evaluation Periods." *Quarterly Journal of Economics*, 112, 631–645.

Heckman, J. J., J. Stixrud, and S. Urzua (2006). "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behaviour." *Journal of Labor Economics*, 24, 411–482.

Heckman, J. J., S. H. Moon, R. Pinto, P. A. Savelyev, and A. Q. Yavitz (2010). "The Rate of Return to the High Scope Perry Preschool Program." *Journal of Public Economics*, 94, 114–128.

Heckman, J. J., S. H. Moon, and R. Pinto (2014). "The Effects of Early Intervention on Abilities and Social Outcomes: Evidence from the Carolina Abecedarian Study." *Unpublished manuscript.*.

Kautz, T., J. J. Heckman, R. Diris, B. ter Weel, and L. Borghans (2014). "Fostering and Measuring Skills: Improving Cognitive and Non-cognitive Skills to Promote Lifetime Success." NBER Working Paper 20749.

Khachatryan, K., A. Dreber, E. Von Essen, and E. Ranehill (2015). "Gender and Preferences at a Young Age: Evidence from Armenia." *Journal of Economic Behavior and Organization*, 118, 318–332.

Lee, S., M. Niederle, and N. Kang (2014). "Do Single-Sex Schools Make Girls More Competitive?" *Economics Letters*, 124, 474–477.

Maddie, S. R., M. D. Matthews, D. R. Kelly, B. Villarreal, and M. White (2012). "The Role of Hardiness and Grit in Predicting Performance and Retention of USMA Cadets." *Military Psychology*, 24, 19–28.

Niederle, M. and L. Vesterlund (2007). "Do Women Shy Away from Competition? Do Men Compete Too Much." *Quarterly Journal of Economics*, 122, 1067–1101.

Niederle, M. and L. Vesterlund (2011). "Gender and Competition." *Annual Review of Economics*, 3, 601–630.

Niederle, M., C. Segal, and L. Vesterlund (2013). “How Costly is Diversity? Affirmative Action in Light of Gender Differences in Competitiveness.” *Management Science*, 59, 1–16.

OECD (2013). *PISA 2012 Results: Ready to Learn: Students’ Engagement, Drive and Self-Beliefs*, Vol. III. PISA, OECD Publishing. http://dx.doi.org/10.1787/9789264201170-en.

Petrie, R. and C. Segal (2014). “Gender Differences in Competitiveness: The Role of Prizes.” Available at SSRN 2520052.

Raven, J., J. C. Raven, and J. H. Court (2004). *Manual for Raven’s Progressive Matrices and Vocabulary Scales*. Harcourt Assessment, San Antonio, TX.

Reuben, E., P. Sapienza, and L. Zingales (2015). “Taste for Competition and the Gender Gap among Young Business Professionals.” National Bureau of Economic Research (No. w21695).

Sutter, M. and D. Glätzle-Rützler (2015). “Gender Differences in the Willingness to Compete Emerge Early in Life and Persist.” *Management Science*, 61, 2339–2354.

Sutter, M., D. Glätzle-Rützler, L. Balafoutas, and S. Czermak (2016). “Canceling Out Early Age Gender Differences in Competition—An Analysis of Policy Interventions”, *Experimental Economics*, 19, 412–432.

Wozniak, D., W. T. Harbaugh, and U. Mayr (2014). “The Menstrual Cycle and Performance Feedback Alter Gender Differences in Competitive Choices.” *Journal of Labor Economics*, 32, 161–198.

Supplementary Data

Supplementary data are available at *JEEA* online.